A MOLECULAR STUDY OF THE CITRUS GENOME THROUGH RESTRICTION FRAGMENT LENGTH POLYMORPHISM AND ISOZYME MAPPING

By

PAN-CHI LIOU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1990

ACKNOWLEDGEMENTS

The author thanks Dr. Fred G. Gmitter, adviser and cochairman, and Dr. Gloria A. Moore, chairman, for their thoughtful supervision and tireless guidance during the course of this study. Thanks also go to Dr. Christine D. Chase, Dr. Paul M. Lyrene, Dr. Jude W. Grosser, and Dr. Ken S. Derrick for serving on the supervisory committee and offering helpful guidance in the planning and conducting of this research. Appreciation is also extented to Dr. Ken C. Cline for advice and allowing the use of his laboratory and equipment. Many thanks are extended to Rick Durham, Shu Huang, Sue Lawrence, Colette Jacano, and Margie Wendell for their friendship and technical asistance. Also, many thanks are extended to the faculty of the Fruit Crops Department and the Citrus Research and Education Center, University of Florida, for the many ways in which they have helped.

Financial support provided by the Council of Agriculture, R. O. C., and asistance from the Taiwan Agricultural Research Institute that made this research possible are also acknowledged.

Finally, the author thanks his wife, Li-hwa, for her unselfish support, typing, and especially love. Sincere thanks go to his parents, the late Mrs. Hsiu-ying Liou who died just before the author's going abroad, and Mr. Hsing-chyuan Liou, for their unconditional love and blessing on his studying abroad.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

A MOLECULAR STUDY OF THE CITRUS GENOME THROUGH RESTRICTION FRAGMENT LENGTH POLYMORPHISM AND ISOZYME MAPPING

By

Pan-Chi Liou

August, 1990

Chairman: Gloria A. Moore
Cochairman: Fred G. Gmitter, Jr.
Major Department: Horticultural Science (Fruit Crops)

Citrus genetic study and cultivar improvement have been difficult via conventional genetic experiments and breeding techniques. Alternative approaches that can enhance efficiency both in genetics and breeding, e.g. studies based on the knowledge of genetic linkages, are a critical need. This study was undertaken to initiate linkage mapping in *Citrus* using RFLP and isozyme analysis.

For *Citrus* RFLP studies, methods that allowed the isolation of high quality and nuclear-enriched DNA suitable for recombinant DNA manipulation were established. A PstI derived *Citrus* genomic library was subsequently constructed to create DNA clones for an RFLP survey; and a quick procedure was developed to facilitate library screening for suitable clones. The overall strategies followed were found helpful in good elimination of organelle DNA; increasing the percentage of single copy clones in the library; and a quick overall survey of a newly constructed library.

From the results of the DNA molecular study, it was concluded that a substantial portion of *Citrus* genome is composed of methylated DNA. The nonmethylated portion of the genome was estimated to contain 73%, 18.9%, and 8.1% single copy, multiple copy, and repetitive sequences, respectively, as determined by high stringency washing conditions in Southern analysis. Some *Citrus* genes may have evolved into multigene families according to the banding patterns revealed by different stringency conditions in Southern analysis.

Using *Citrus reticulata* cv. Clementine, *C. paradisi* cv. Duncan, and a hybrid derivrd from these two cultivars (LB 1-21) as a screening filter, it was found that insertion/deletion is the major mechanism causing RFLPs. A DNA clone was identified that was suggestive of a transposon-related sequence in *Citrus*. Moreover, some unusual data in the progeny analyses strongly suggested the mechanism of reversional excisions in the *Citrus* genome. A hypothesis of transposon activity was proposed based on these observations and analyses.

Forty-two markers (36 RFLP clones and 6 isozymes) showed segregations in a population of 65 seedlings derived from the backcross of LB 1-21 x 'Clementine' mandarin, and were analyzed for the linkage survey. A linkage map including eight linkage groups, totally comprising 35 mapped loci (29 RFLP and 6 isozyme loci), was established.

CHAPTER 1

INTRODUCTION

Citrus has been cultivated for several thousand years (Swingle and Reece, 1967). Today it is an important fruit crop throughout the tropical and subtropical regions of the world. It has been predicted that *Citrus* will remain a major fruit crop and an important part of diets in a vitamin-conscious, rapidly developing modern world (Burke, 1967). In the U.S., systematic *Citrus* breeding studies were begun in the late 19th and early 20th centuries. Later, other breeding studies were carried out in Java, Italy, the Philippines, Japan, and the U.S.S.R. (Cameron and Soost, 1984). However, despite active and vigorous *Citrus* breeding programs, very few cultivars have been produced via standard breeding methods (Hearn,1973; Gmitter, 1985). Most commonly grown cultivars have been derived from bud mutations or well-adapted superior seedlings (Hodgson, 1967). Several factors have made *Citrus* breeding a slow and difficult process. These include a long period of juvenility, great heterozygosity, the quantitative inheritance of most characters, and large individual plant size. Also, self-incompatibility and polyembryony are present in many varieties and make cultivar improvement more difficult (Soost and Cameron, 1975). These factors have also hindered *Citrus* genetic studies and resulted in a poor understanding of *Citrus* genetics. Thus, advances in *Citrus* breeding and genetics by conventional methods will continue to be slow (Cameron and Soost, 1984).

Recently molecular biological technology has provided new approaches for plant breeding. A restriction fragment length polymorphism (RFLP) linkage map can be constructed for a crop of interest, to provide a method for selecting or tagging desirable genes via their linkage to easily detectable RFLP markers (Tanksley et al., 1989). Plant variety improvement can also be enhanced through the integration of RFLP techniques into conventional plant breeding schemes. This technique has several invaluable advantages over mapping with morphological markers, as will be discussed in the following chapter, and it is becoming a powerful tool for plant genetic and breeding studies in many species (Helentjaris et al., 1985, 1986; Tanksley et al., 1989).

Little information on *Citrus* molecular biology or genetic linkage has been published (Torres et al., 1985; Roose, 1988). The overall objective of the studies described in this dissertation was to initate a molecular genetic study of *Citrus* through DNA cloning, genome characterization and linkage map construction. The first chapter to follow is a review of literature pertinent to genetic mapping techniques in general, RFLP mapping techniques in particular, and *Citrus* genetics. The three chapters following the literature review will deal respectively with research into genomic library construction and clone screening, genome characterization by RFLP analysis, and gene mapping using RFLP and isozyme analysis. Important points and conclusions that are particularly relevant to *Citrus* genetics will be discussed within the individual chapters. The final chapter will summarize the conclusions, and describe a simplified method for enhancing RFLP studies of plants, especially those species for which molecular studies have not yet been initiated. Also, some assumptions made in the two previous chapters will be integrated into a hypothetical model of transposon activity for *Citrus* genetics.

CHAPTER 2

REVIEW OF THE LITERATURE

Gene Mapping

A genetic map describes the linkage relationships among gene loci. It indicates the order of the loci and recombination frequencies (map units) between them. Gene maps are usually depicted as a set of lines with gene loci located along the lines.

History of Gene Mapping

The history of gene mapping can be dated back to 1911 when Sturtevant realized that linkage information could be used to determine the relative position of genes along a chromosome (Suzuki et al., 1981; Donis-Keller et al., 1987). Genetic maps of numerous organisms have been published since that time, including animals, plants, and relatively recent maps of viral and bacterial chromosomes (Goodenough, 1984). The advancement of gene mapping capabilities is mainly based on two important factors, the use of new molecular genetic markers, and the development of techniques for linkage analysis.

During the early years of gene mapping studies, only morphological markers under single gene control could be mapped (Landry et al., 1987; Tanksley et al., 1989); and only experimental populations, such as F2 or backcross progeny, were useful because of the statistical considerations involved in mapping (Fisher and Balmakund, 1928; Mather, 1938). Therefore, historically only those species with

numerous visible mutants and available tester lines could be used for defining gene maps (Lander et al., 1987). Notable examples are *Drosophila*, *Zea mays*, *Lycopersicon esculentum*, and *Hordeum vulgare* (Immer and Henderson, 1943; McMillin, 1983; Lander et al., 1987).

However, many limitations to genetic mapping studies have been minimized by the development of new technologies. The use of isozyme markers for mapping began in the late 1960s (McMillin, 1983). A more powerful technique, RFLP analysis, was employed during the last decade for mapping. These new types of markers, especially RFLPs, have greatly impacted and facilitated the construction of gene maps for almost every species (Tanksley et al., 1989).

In addition to the use of new types of markers, the development of more exact and adaptable methods for linkage analysis progressed greatly during the 1930s and 1950s (Mather, 1938; Allard, 1956). The application of computer programs to linkage analyses began in the 1970s (Ott, 1973). All of these accumulative efforts have greatly increased capabilities to create useful and valid gene maps for any organisms of interest.

Techniques Developed for Mapping

Recomination-frequency mapping. The most important and common method of gene mapping uses recombination frequency to determine the relative distance between two linked genes (one percent recombination is defined as one centimorgan, Rieger et al., 1976). This method is based on the theories of probability and was established, in large part, using *Drosophila melanogester* as the experimental organism (Goodenough, 1984). The principle of rough proportionality, postulated by Sturtevant in 1911, states that the frequency of crossover between two

linked genes is roughly proportional to the physical distance between them. This principle remains the basis for genetic mapping studies today (Suzuki et al., 1981).

The measurement of linkage is quite simple and straightforward with complete data, and no statistical methods are necessary (Fisher and Balmukand, 1928). For instance, with the *Drosophila* data used by Sturtevant for his mapping studies (i.e. data from the "double heterozygote x double recessive" progeny), the calculation of the proportion of the "recombination" classes to the total count provided satisfactory recombination frequencies, and thus genetic distances (Fisher and Balmukand, 1928; Suzuki et al., 1981). However, in most cases, e.g. when using F2 progeny, the classification data cannot provide the linkage information in a straightforward manner; therefore, statistical methods are necessary to examine the data for evidence of linkage. In such situations, three different statistical analyses are necessary to create an accurate map (Mather, 1938). First, the segregation ratios at each locus need to be tested for conformation with expected Mendelian inheritance ratios. Second, an analysis for independent assortment between jointly segregating loci is performed to identify linked gene pairs. Finally, "the estimation of linkage" or "the estimation of recombination intensity" is made to locate genes (Mather, 1938; Suiter et al., 1983).

In general, chi-square and contingency chi-square tests are satisfactory, respectively, for the first and second analyses (Mather, 1938). However, the estimation of linkage, which bears no relation to the calculation of chi-square, can be quite complicated or confusing in many cases based on the properties of the population used for producing the segregating data (Fisher and Balmukand, 1928; Immer, 1930; Mather, 1938). During the early 1900s, many different methods were developed and used to estimate linkages, including the product moment method, additive method, maximum likelihood method, weighted mean method, and

minimum chi-square method (Fisher and Balmukand, 1928). In some cases, two or more methods applied to the same data gave widely differing results (Immer, 1930). Geneticists and statisticians realized during the 1920s that a general method capable of satisfying all ordinary linkage problems was necessary, and the most useful method should be easy to use and have a probable error as small as possible (Immer, 1930).

Fisher and Balmukand (1928) first concluded that the most useful methods for linkage estimation were the product moment method and the method of maximum likelihood. Although the product moment method can only be applied to populations in which four-class segregations occur (Fisher and Balmukand, 1928; Allard, 1956), its simplicity of application has made it the best method when using F2 populations for linkage studies (Fisher and Balmukand, 1928; Immer, 1930; Stevens, 1939; Immer and Henderson, 1943). Simplicity of method was extremely important during that time when tedious and laborious calculation by hand was always a limiting factor for practical linkage studies. However, Mather (1938) concluded that the method of maximum likelihood was the only method that can lead to efficient estimates for all types of problems. When different experimental designs and complex data are encountered, it is the only method giving satisfactory results in linkage studies.

In the method of maximum likelihood, "likelihood" is the probability that each meiosis under study would come out exactly the same if the experiment were repeated (Lander et al., 1987). This method requires the computation of the probability that phenotypes turn out to be as actually observed by considering all possible outcomes in computation (Ott, 1974). The purpose of this mathematical process is to establish a function of likelihood by using the multinomial expansion and logarithm method, and to determine a value of the variable that maximizes

likelihood, to be used as the best estimate of the parameter (recombination frequency) (Mather, 1938). Its basic principle is that the value of likelihood largely depends on the recombination frequency (Ott, 1974).

The calculations of likelihood are tedious and time consuming, especially in progenies where numerous loci are segregating simultaneously (Suiter et al., 1983). First, formulae for computing maximum likelihood and standard errors were presented by Immer in 1934 (c.f. Kramer and Burnham, 1947). Mather (1938) gave a complete discussion of the use of the maximum likelihood method for linkage estimation, including the planning of experiments and the disturbed segregations that may occur in many cases. Immer and Henderson (1943) presented tables that greatly reduced the necessary calculations of this method. In 1946, Fisher (c.f. Kramer and Burnham, 1947) developed a simple method of scoring linkage data based on maximum likelihood formulae, that included estimates of the significance of deviations from independence. Kramer and Burnham (1947) made their contribution by modifying the existing formulae to accomplish different specific purposes and by summarizing the simplest methods available for calculating maximum likelihood. Morton (1955) computed and tabulated likelihoods in the form of so-called LOD-scores, defined as the logarithm of the relative likelihood of linkage versus nonlinkage, which simplified use of the maximum likelihood method. Morton's LOD-score method is useful when using some natural populations for linkage studies, such as in human gene mapping (Ott, 1974). Allard (1956) provided complete equations and tables for various F2, F3, and backcross situations, including equations for a complementary or duplicate pair, a gene pair producing a 13:3 ratio, and a third gene pair.

The accumulated work of these geneticists and statisticians greatly facilitated application of the maximum likelihood method. The well-developed techniques of

the maximum likelihood method have proven invaluable for gene mapping, for they have reduced the limitations of experimental planning. This is crucial for species in which inbred lines are not available or matings cannot be arranged (e.g. trees or humans) (Lander et al., 1987). Thus, the efficiency of gene mapping in many organisms was enhanced, especially when computerized calculations (see next section) and new genetic markers (e.g. RFLPs) were used for mapping (Lander et al., 1987).

Though the construction of linkage maps can be facilitated by using computers with a maximum likelihood program, maps derived from natural populations usually are not as precise as from experimental populations. Frequently data from natural populations are fundamentally incomplete. Population size and the number of co-informative loci are usually not under control (Lander et al., 1987). New methods, again based on the maximum likelihood, have been developed to overcome this problem. The probability that a possible map would have given rise to the observed data is computed; this is called "the likelihood of the map". A good map is the one with the highest likelihood. By searching over many possible maps, one can find the map with maximum likelihood (Lander et al., 1987).

Another problem also needs to be addressed; i.e., the traditional method for inferring gene order by two-point crosses cannot always give reliable results when using incomplete data derived from natural populations. Multipoint linkage analysis instead of two-point analysis is required to overcome this problem. The same type of modification, called multi-locus analysis, is necessary also to compute the likelihood of a given map (Lander et al., 1987). Mathematically, the multipoint linkage analysis (from experimental data to a constructed map) and the multilocus analysis (from possible maps to observed data) are the same technique, though conceptually performed in opposite directions.

Elston and Stewart (1971) provided the first general algorithm for computation of the likelihood of a given map. The Elston-Stewart algorithm was good only for two-locus analysis; the computation time grew exponentially with the number of loci in the map when the algorithm was used for multilocus analysis, even on a supercomputer. When the number of loci was large, as it must be to create a reasonably useful map, the algorithm became impractical (Morton et el., 1986). To solve this problem, Lathrop et al. (1986) described a modified Elston-Stewart algorithm that resulted in a substantial increase in computation speed and made multilocus analysis easier to perform. Lander and Green (c.f. Lander et al., 1987) provided a different algorithm for high speed likelihood computation that has been proven to scale linearly, rather than exponentially, with increasing locus number.

Computer programs for gene mapping. Several computer programs are available for linkage study, including 2 in FORTRAN (LIPED [Ott, 1973]; and LINKAGE [Lathrop and Lalouel, 1984]), LINKAGE-1 in PASCAL (Suiter et al., 1983), and MAPMAKER in C programming language (Lander et al., 1987). These programs are based on the maximum likelihood method, although they use different algorithms for computation.

LIPED was the first computer program used in linkage studies. It was designed in 1973 to calculate the likelihood of some specific types of human pedigrees, e.g. those without consanguineous marriages. An extension to general pedigrees was implemented in a new version in 1976 (Ott, 1973, 1974, 1976). This program used the Elston-Stewart algorithm to compute likelihood and Morton's LOD-scores for linkage estimation.

LINKAGE-1 was designed for linkage analysis using a variety of segregating progenies, e.g. F2 and backcross types (Suiter et al., 1983). The program accepts

dominant and co-dominant genes as segregating loci. A chi-square test is used for checking the single-factor segregation, and a contingency chi-square test for linkage detection. For each pair of linked loci, the recombination frequencies and standard errors are calculated. The maximum likelihood method used in the linkage estimation is based on the formulae provided by Mather (1938) and Allard (1956).

Both the LIPED and LINKAGE-1 programs are designed only for two-point analysis (Suiter et al., 1983; Lander et al., 1987). Using these programs, a linkage map must be constructed by hand based on the two-point information. Lander et al. (1987) have reported that a map deduced from such data can yield incorrect gene orders, especially when a limited number of co-informative meioses are studied. Other programs designed to perform multipoint analyses can help overcome this problem. LINKAGE was the first program capable of doing multipoint analysis. It uses the modified Elston-Stewart algorithm to calculate likelihood (Lander et al., 1987). This program can also be used for analyzing quantitative genetic data, and is suitable for a wide range of applications in linkage studies (Lathrop and Lalouel, 1984; Lathrop et al., 1985; Barker et al., 1987).

MAPMAKER was specially designed to use multipoint linkage analysis for construction of accurate genetic maps. It has been especially useful for the construction of primary genetic linkage maps using RFLP data, either from F2 experimental populations or from two- and three-generation nuclear families in natural populations (Lander et al., 1987; Donis-Keller et al., 1987). The program uses the Lander-Green algorithm for high speed computation of likelihood, which makes it practical to study large numbers of loci simultaneously. The authors have also developed systematic strategies that can be used to construct detailed genetic linkage maps when using this program.

Other techniques for mapping. There are several other techniques that can be used in gene mapping. *In situ* hybridization, using labeled probes of cloned DNA, can provide chromosome assignments for a locus and its linkage group. This technique has been successfully used in humans, mice, chickens, *Drosophila*, and wheat (Goodenough, 1984; Rayburn and Gill, 1985). Lichter et al. (1990) reported that the use of chromosomal *in situ* suppression (CISS) hybridization in conjunction with fluorescent detection of hybridized probes can rapidly and precisely map a large number of cloned genomic DNA segements to human chromosomes. This high-resolution mapping method is also capable of determining the gene order along the chromosome. Gene mapping by use of somatic hybrid cells has been reported in mammals using hybrid cell clones (Klug and Cummings, 1986). Mutations resulting from DNA deletions are sometimes useful for mapping a locus by detection of trait deficiency. This technique, called deficiency or deletion mapping, is used frequently to map viral genomes (Goodenough, 1984). Aneuploids, including trisomics and monosomics, can be used to assign genes to chromosomes. The determination of which chromosome carries a particular genetic locus depends upon allele dosage effects of isozymes in primary aneuploid plants or distorted segregation in backcross or F2 progeny, which occurs when the chromosome with the relevant locus is in either the monosomic or the trisomic condition. For a few plant species, mapping has been facilitated by using these kinds of approaches, e.g. maize, barley, tomato, wheat, and pepper (McMillin et al., 1979; Fobes, 1980; Moore and Collins, 1983; Tanksley, 1984).

Applications of Gene Maps

A well defined gene map, consisting of easily scored polymorphic marker loci spaced throughout a genome, is a fundamental tool for genetic studies. The

understanding of genetic mechanisms controlling traits of economic importance can be enhanced by using the information provided by a linkage map (Tanksley, 1983; Beckmann and Soller, 1986; Lander et al., 1987; Landry et al., 1987). Such a map could be an invaluable aid to breeders as well, because it enables them to determine the likelihood of synthesizing genotypes with desired combinations of beneficial alleles. Without gene maps, this will be a long and difficult procedure (McMillin, 1983; Tanksley et al., 1989).

In a detailed linkage map, markers that are tightly linked to desirable genes may be discovered. Such markers could then be used to easily detect the desirable gene in segregating populations (Moore and Collins, 1983). A classic example can be found in the tomato linkage map, where the acid phosphatase locus (*Asp-1* allele) is tightly linked to a gene for nematode resistance (*Mi* allele). This linkage enables breeders to select for resistance at early seedling stages by isozyme analysis instead of by exposure to the parasite. Selection at the seedling stage for zymotype has enhanced the efficiency of the nematode resistance breeding programs (Bolkan et al., 1987).

RFLP markers potentially are very powerful tools for tagging disease resistance genes. Examples have been given by tomato and human linkage maps. In tomato, a RFLP marker linked to the *Fusarium oxysporum* resistance gene *I2* has been found, which enables efficient screening without resorting to test inoculation with the pathogen (Sarfatti et al., 1989). In humans, a DNA segment encoding two marker genes very tightly linked to Huntington's disease has been found (Gilliam et al., 1987). Tightly linked RFLP clones may also function as starting points for various molecular approaches, such as chromosomal walks aimed at direct cloning of genes of interest (Sarfatti et al., 1989; Tanksley et al., 1989).

The linkage between quantitative trait loci (QTL) and monogenic loci can be detected statistically from calculations based on the normal distribution (Tanksley et al., 1982). The mapping of QTL has been greatly enhanced using the complete linkage map of restriction fragment length polymorphism (Paterson et al., 1988). In this manner, some QTLs in tomato have been investigated and mapped (Tanksley et al., 1982; Zamir et al., 1982; Osborn et al., 1987; Tanksley and Hewitt, 1988).

Genetic Markers Used for Mapping

Morphological markers. Morphological markers that provide easily recognizable polymorphisms, such as differences in color, shape and height, etc., were the first type of markers used for gene mapping (Landry et al., 1987). Ease of scoring genotypes and low cost are the major advantages for using this type of marker. There are several disadvantages, however. Morphological markers can be influenced by environmental factors, or epistatic interactions may occur. When allelic dominance or recessiveness plays a role in gene expression, only phenotypes can be detected (Goodenough, 1984; Klug and Cummings, 1986). Detailed linkage maps require many morphological mutants, but this condition is met by very few species, e.g. barley, maize, tomato (Immer and Henderson, 1943; Tanksley, 1983). Moreover, it is rare that several markers are found in the same test lines, even though many mutations are available within the species. These limitations make experimental planning and execution of mapping studies laborous; a large number of segregating populations is required to develop a linkage map because only a limited number of loci co-segregate in each cross (Landry et al., 1987).

Biochemical markers. There are several different types of biochemical markers. For example, polymorphisms for CTV resistance detected in *Citrus* by Enzyme-Linked Immuosorbent Assay (ELISA) (Yoshida, 1985), and the different

genotypes in *Citrus* detected by polyphenol oxidase-catalyzed browning of young shoot extracts (Esen and Scora, 1975), were based on biochemical reactions of gene products. However, isozymes are the best documented and most useful biochemical markers for genetic studies and gene mapping.

The term "isozyme" was coined to describe the different molecular forms of proteins that exhibit the same enzymatic specificity, and can be studied through gel electrophoresis techniques (Goodenough, 1984). Isozyme technology was developed first in the 1950s and used for gene mapping of *Triticum aestivum* in 1969 (Brewer et al., 1969; McMillin, 1983). Most isozyme alleles are co-dominant in nature thus allowing direct study of genotypes, and they usually follow Mendelian inheritance. Isozyme epistatic interactions are rare, and in most cases are not tissue or developmental stage specific. Therefore screening plants at an early stage is possible, and any plant tissue can be sampled. The technique is relatively inexpensive, not time consuming, and the process is minimally destructive because only small amounts of tissue are needed. These characteristics have made isozyme analysis very versatile (Tanksley and Rick, 1980; Ben-Hayyim et al., 1982; Adams, 1983; Moore and Collins, 1983). In addition, the efficiency of linkage studies can be significantly enhanced by using a single or very few experimental populations derived from parents that have been selected on the basis of many segregating markers (Landry et al., 1987). Isozymes provided an opportunity to remove the limitations to linkage map construction in organisms that lack morphological markers, and thus significantly impacted gene mapping research (Tanksley, 1983).

Molecular markers. Molecular markers allow direct detection of polymorphism at the DNA molecular level. RFLPs are the most useful molecular markers for gene mapping. The term "restriction fragment length polymorphism" (RFLP) was coined to denote the differences in molecular weight of homologous

DNA fragments created by restriction endonucleases (Botstein et al., 1980; Hudson, 1982). RFLPs can be caused by base pair changes, DNA rearrangement, or insertion/deletion events, and they can provide information on nucleotide diversity and genetic variation within/among populations (Helentjaris, 1987; Hudson, 1989). This new type of marker was first proposed for human gene mapping (Botstein et al., 1980), and has recently been used in plant gene mapping studies (Burr et al., 1983; Soller and Beckmann, 1983; Helentjaris et al., 1985; Lander et al., 1987; Tanksley et al., 1989).

In genomes of small size, such as organelles, RFLP analysis is easy to perform. A limited number of restriction fragments are produced after treatment with a restriction enzyme and discrete bands can be observed after gel electrophoresis, so that a restriction map can be constructed (Avise et al., 1979; Seyer et al., 1981; Medgyesy et al., 1985; Kaneko et al., 1986; Rines et al., 1988). However, RFLP analysis of eukaryotic nuclear genomes is not so straightforward, because when the genomic DNA of an eukaryote is cut with a restriction endonuclease, millions of DNA fragments in a continuous range of sizes can be produced. These fragments can not be directly studied through gel electrophoresis. More complex techniques have to be employed for solving this problem. The basic principle is that a labeled probe is needed to visualize the polymorphic fragments. Cloned DNA is required for RFLP analysis of eukaryotes as a probe for Southern analysis to resolve fragments and detect RFLPs (Old and Primrose, 1985; Watson et al., 1983, 1987).

RFLPs share many of the advantages of isozymes for use as genetic markers. Nuclear genomic RFLPs are co-dominant and inherited in a simple Mendelian genetic fashion (Helentjaris, 1987) revealing genotypes in Southern analysis (Botstein et al., 1980). No epistatic interactions occur with RFLPs (Beckmann and Soller,

1986), and they are not tissue or developmental stage specific (Beckmann and Soller, 1986). Although techniques for mapping RFLPs are more complicated and difficult than those required for isozyme analysis, RFLP markers have a number of advantages over isozymes and other genetic markers. RFLP analysis is done at the DNA level. It is not necessary to go to the gene product level, therefore RFLP analysis does not require any knowledge of the biochemical nature of a trait. Thus, RFLPs can be useful for mapping cases where DNA sequence changes do not alter the amino acid sequences of gene products, or where DNA changes defeat gene functions (Botstein et al., 1980). The potential number of RFLPs in an organism is virtually unlimited, since on average at least 10^6 fragments can be created by an endonuclease and over 100 endonucleases are available to date (Helentjaris, 1987; Lander et al., 1987). Once a source of probes is available, a virtually infinite number of loci can be examined using different probe/enzyme combinations, and an abundance of genetic markers can be provided (Lander et al., 1987). Thus, RFLPs have made it possible to develop detailed genetic maps using only one or a few populations (Landry et al., 1987). This is extremely useful for species such as tree crops and humans, where co-informative loci are rare or where tester lines as well as experimental populations are difficult to establish.

Some genetic linkage maps based on RFLP markers have been published including human (White et al., 1985; Donis-keller et al., 1987); maize (Helentjaris, 1987), lettuce (Landry et al., 1987), tomato (Helentjaris et al., 1986), potato (Tanksley et al., 1989; Bonierbale et al., 1988), wheat (Kam-Morgan and Gill, 1989), *Arabidopsis thaliana* (Chang et al., 1988; Meyerowitz, 1989), and the fungus *Bremia lactucea* (Hulbert et al., 1988). It is anticipated that in the near future genetic linkage maps based on RFLPs will be available for many organisms of interest (Lander et al., 1987; Tanksley et al., 1989).

Molecular Techniques Used for RFLP Studies of Plants

Introduction

RFLP analysis is based on the use of cloned chromosomal DNA fragments as genetic probes. The technology consists of four essential parts: the generation of DNA fragments, DNA cloning, library construction, and the detection of polymorphism by Southern analysis. When RFLP clones are derived from a genomic library (see following sections), DNA isolation is also a crucial step. Most of the necessary techniques that have become powerful tools for genetic analysis in the past decade were developed during the past 20 years.

DNA Isolation

Special problems encountered in plant DNA isolation. For most molecular genetic studies, DNA of high molecular weight and purity is required. Good quality DNA is necessary for digestion with the restriction endonucleases and for cloning (Rogers and Bendich, 1988). It is more difficult to isolate DNA of such quality from plants than from other organisms. The first problem encountered is disruption of the rigid cell wall unique to plants. The methods necessary for cell disruption and DNA extraction may shear the native DNA molecule, thus, increasing the DNA yield often decreases the length of recovered fragments. Many purification procedures include vigorous mixing of native DNA with deproteinizing agents that can result in additional shearing of the DNA (Ohyama, 1975; Murray and Thompson, 1980). During these procedures, released DNA may be attacked by endogeneous nucleases. Second, polysaccharides and tannins are often a major problem in plant DNA isolation because they are difficult to separate from DNA (Murray and Thompson, 1980). The identity of the exact contaminant that co-purifies with plant

DNA is still unknown. It does not appear to be a non-charged polysaccharide such as starch or pectin; rather, it may be a charged polymer, possibly similar to heparin (Slightom and Drong, 1988). Finally, the success of any particular DNA isolation procedure depends upon the plant group being studied. A protocol that works with one plant group will often fail miserably with others. All these factors have made plant DNA isolation a difficult project, especially when initiating a study for a new species (Slightom and Drong, 1988; Doyle et al., 1990).

Methods of DNA isolation. Many procedures and strategies have been developed for plant DNA isolation that may be tested for adaptation to a new species. Most procedures consist of four main steps: cell disruption, DNA-protein complex dissociation, removal of protein, and further purification and condensation.

Plant cell walls are disrupted by grinding tissue frozen with liquid nitrogen. Frozen tissue is prefered because it can be efficiently disrupted while the DNA is unhydrated and less susceptible to shearing (Holl, 1975; Murray and Thompson, 1980; Dellaporta et al., 1983; Doyle et al., 1990). The liquid nitrogen procedure usually yields DNA of higher average molecular weight than methods where fresh tissue is ground directly in buffer (Doyle et al., 1990). Freshly-formed protoplasts of some species, e.g. *Daucus carota* and *Ammi visnaga*, have been used to avoid the mechanical milling (Ohyama et al., 1972; Ohyama, 1975). Isolated intact nuclei may be a good source for isolation of long plant DNA fragments free of cytoplasmic DNA; however DNA yields are often very low, and the techniques required are elaborate and involved (Luthe and Quatrano, 1980; Murray and Thompson, 1980; Mathew, 1984; Keim, 1987; Paul et al., 1987). Mederic et al. (1987) treated plant tissues with glycerol to avoid mechanical disruption of cells and shearing damage to DNA.

Once cell walls are broken, a detergent such as SDS (sodium dodecyl sulfate) is used, usually in company with an appropriate lysis buffer, for lysis and dissociation of the DNA-protein complex (Mathew, 1984). Several different detergents have been tested for the elimination of polymeric contaminants, and some such as CTAB (cetyltrimethyl ammonium bromide) and sarkosyl were found to be promising (Mettler, 1987; Slightom and Drong, 1988). The CTAB method is preferable because of its simplicity and quickness (Murray and Thompson, 1980; Rodgers and Bendich, 1988; Slightom and Drong, 1988). In the CTAB method, DNA can be selectively precipitated as a CTAB-DNA complex through the adjustment of NaCl concentration, and therefore separated from polysaccharides. Thus the polymeric contaminants can be greatly reduced. (Murray and Thompson, 1980).

After the detergent treatment, many different protocols remove proteins and other compounds by extraction with phenol/chloroform (with isoamyl alcohol, 24:1). Proteinase K may optionally be used to degrade proteins (Maniatis et al., 1982; Mettler, 1987; Doyle,et al., 1990). CsCl gradients for DNA purification are still used by some researchers; although this method is very efficient, it is time consuming and expensive (Murray and Thompson, 1980; Keim, 1987; Doyle et al., 1990).

Finally, purified DNA is condensed by ethanol or isopropanol precipitation, sometimes in company with ammonium acetate. This step also removes salts or reaction products (Maniatis et al., 1982; Crouse and Amorese, 1987; Mettler, 1987; Doyle et al., 1990).

Recombinant DNA Techniques

The term "recombinant DNA" denotes a DNA molecule in which sequences of different origins have been combined by *in vitro* manipulations (Oliver and Ward, 1987). Recombinant DNA techniques comprise two essential parts: the method for generating DNA fragments, and the reactions that join a DNA fragment (insert) to the vector.

DNA fragment generation for cloning. There are two sources from which DNA fragments can be generated for insertion in the vector to study the nuclear genome. One source is genomic DNA, which is isolated from tissue as described in the last section; the other source is cDNA, which is synthesized from mRNA.

Purified DNA digested with a restriction endonuclease and fractionated either by sucrose gradients or gel electrophoresis is most commonly used to generate genomic DNA fragments (Christiansen, 1984; Perbal, 1984; Davis et al., 1986). Restriction endonucleases that produce cohesive termini (sticky ends), such as *Eco*RI and *Pst*I, are preferably used to create termini on the insert and vector to facilitate the recombination between the two (Sambrook et al., 1989). In some cases, mechanical shearing followed by addition of linker molecules or homopolymer tailing sequences on both ends have been used to produce DNA fragments (Sambrook et al., 1989).

A cDNA is a complementary copy of mRNA sythesized *in vitro* in the presence of the enzyme reverse transcriptase, that functions by creating a single DNA strand upon an RNA template. (Watson et al; 1983). The newly synthesized DNA sequence is made double stranded through a series of steps in the reaction. Finally, a homopolymeric tailing sequence or linker molecule is added to each end of the duplex cDNA fragment for ligation. Sometimes blunt-end ligation, without

linkers or homopolymer tails, is performed directly (Maniatis et al., 1982; Watson et al., 1983).

Cloning vectors. A vector, also called a cloning vehicle, is a DNA molecule derived from a plasmid or bacteriophage into which fragments of DNA may be inserted. The procedure of inserting a particular gene or other DNA sequence into a vector is referred as "cloning" or "(to) clone" (Old and Primrose, 1985; Oliver and Ward. 1987). The vector should contain one or more unique restriction sites for the convenience of the cloning process, and should be capable of autonomous replication in a defined host bacterium so that the cloned DNA is reproduced (Maniatis et al., 1982).

Plasmid derived- and phage derived-vectors are the two basic groups of cloning vectors. These groups differ in some characteristics. In the plasmid group, vectors are inherited and reproduced in an extra-chromosomal state. DNAs are cloned through insertion, and cloned vectors are transformed into cells of the host organism. The limit of insert (DNA to be cloned) size is smaller (usually less than 10 kb) than DNA cloned into phage vectors (Watson et al., 1987). In the group of phage vectors, insert DNAs are cloned through insertion of a specific site or in replacement for the non-essential portion of the vectors. Cloned vectors are packaged and transfected into host cells, and they are reproduced during the lysogenic and lytic cycles after transfection. The cloned vectors need to be in the size of ca 44-52 kb (kilo base pair) for successful packaging into an infective phage particle, so there is positive selection for the insert size (Oliver and Ward, 1987).

Some categories of vectors are constructed to have properties of both phage and plasmid vectors. Examples are cosmids and phagemids. A cosmid is a plasmid vector with a phage *cos* site (cohesive ends of certain phage molecules responsible for forming concatemeric λDNA into viral head particles) and one or more selectable

markers such as a drug resistance gene. Cosmids usually have much larger capacity for harboring insert DNA (up to 52 kb) (Grierson and Covey, 1984). A phagmid is a hybrid molecule formed between a plasmid containing multiple λ phage *att* site (the sequence where integration and excision of phage λ occurs inside the host cell) thus making it capable of replicating as a plasmid (non-lytically) or a phage (lytically), and providing flexibility for more sophisticated genetic manipulations (Oliver and Ward, 1987). Different vector classes have been constructed within every group described above to confer some well-defined phenotypes on the host organism that are either selectable (e.g. drug resistance in plasmid vectors) or readily detected (e.g. plaque formation in phage vectors) (Davis et al., 1986).

<u>The ligation reaction</u>. Once DNA fragments and vector DNA with the same sticky ends are prepared, T4 DNA ligase (an enzyme that catalyses the ATP-dependent formation of a phosphodiester bond between a 5' phosphate and a 3' hydroxyl in duplex DNA) is most commonly used to join the DNA sequences together in recombinant molecules (Maniatis et al., 1982). T4 DNA ligase is also capable of joining two blunt-ended DNA fragments together at high enzyme concentrations. The created recombinant DNA is then used for library construction (Oliver and Ward, 1987).

<u>Host bacterial strains, vectors, and transformation/transfection</u>. In library construction for RFLP studies, all of the host bacterial cells are *E. coli* derived and have specific genetic properties. Different cloning vectors require specific host cells for transformation (plasmid vectors) or transfection (phage vectors), not only for efficiency but also for ease of selection and screening (Maniatis et al., 1982; Hanahan, 1983; Old and Primrose, 1985; Davis et al., 1986). Many different host/vector combinations have been evaluated (Hanahan, 1983; Davis et al., 1986).

In 1970, Mandel and Higa demonstrated that uptake of bacteriophage λ DNA (transfection) was enhanced by treatment of bacterial cells with calcium chloride. Cohen et al. (1973) demonstrated that the same treatment also worked for transforming bacteria with plasmid DNA. Usually transfection is more efficient than transformation (Grierson and Covey, 1984; Old and Primrose, 1985). Many variations of the basic technique have been developed for higher transformation efficiency (Jones et al., 1981; Maniatis et al., 1982). Hanahan (1983) devised a set of conditions for optimal efficiency of transformation (expressed as number of transformants per μg plasmid DNA) applicable to most *E. coli* k12 strains. Typically, efficiencies of 10^7 and 10^8 can be achieved.

Library construction

A gene library, or gene bank, is a collection of recombinant DNA molecules containing inserts (Oliver and Ward, 1987). The gene library may be a genomic library or a cDNA library, depending on the source of DNA fragments used for cloning. The main difference between these two kinds of library is that a cDNA library will only cover the coding sequences of the genome that are expressed as mRNA; even the introns, which often occur within eukaryotic genes but are spliced from the mature mRNA transcripts, are not included in a cDNA library. In contrast, a genomic clone should randomly come from the genome, so a genomic library should more completely represent the whole genome (Watson et al., 1983; Old and Primorose, 1985).

Clone selection from the library

There are two approaches available for selecting DNA clones for RFLP studies, dot blot hybridization and Southern analysis (Southern blot hybridization).

In dot blot hybridization, plasmid clones isolated from bacterial colonies are denatured and transferred onto an appropriate membrane such as nitrocellulose, and then hybridized with labeled nuclear plant DNA (Maniatis et al., 1982; Helentjaris et al., 1985; Landry and Michelmore, 1985). This approach is useful for detecting clones carrying repetitive plant sequences, which will give relatively strong signals after hybridization. However, the usefulness of this approach is very limited because the sensitivity does not allow a distinction between low copy number and single copy sequences, nor does it allow the detection of RFLPs (Landry and Michelmore, 1985).

Southern analysis has been an extremely powerful tool for RFLP studies analyzing eukaryotic genomes (Watson et al., 1983; Klug and Cummings, 1986). It allows detection of the presence of specific DNA sequences and their copy number in the genome. The technique includes the following steps: 1) the fractionation of genomic DNA digests by agarose gel electrophoresis; 2) the denaturation and transfer of DNA from the gel to a nitrocellulose membrane, a process called Southern blotting (the transferred filter is called a blot); 3) a probing process that uses labeled DNA sequences (probes) from a library to hybridize to the blotted DNA; and 4) the detection of restriction fragments homologous to the probe (Southern, 1975). Different modifications of the procedure have been made on the basic technique since it was first developed. Important modifications include: 1) the transfer of DNA from agarose gels to nylon membranes instead of nitrocellulose (Reed and Mann, 1985); 2) improved prehybridization/hybridization procedures (Church and Gilbert, 1984); and 3) rapid transfer of DNA by electrophoretic blotting instead of capillary action (Reed and Mann, 1985).

Genetic Studies in Citrus Over the Past 90 Years

General Aspects

The earliest document of *Citrus* genetics can be traced back to 1878 when Strasburger first described the phenomenon of polyembryony. However, most knowledge of *Citrus* genetics has been gained in the past ninety years (Swingle and Reece, 1967; Cameron and Frost, 1968). In comparison with most other plants, *Citrus* genetics is a difficult subject to study because of the long juvenile period, large tree size, great heterozygosity of most species, and frequent production of nucellar progeny (Soost and Cameron, 1975). Most available genetic information has been obtained as a by-product of variety improvement efforts.

Ploidy and genome size. Citrus chromosomes are small, about 2 μ long at first metaphase of meiosis (Soost and Cameron, 1975). The basic chromosome number in *Citrus* and other genera in the subfamily Aurantioideae is x = 9. Diploidy (2n=2x=18) is the general rule in *Citrus* and its related genera such as *Fortunella* Swingle, *Poncirus* Raf., *Microcitrus* Swingle, *Eremocitrus* Swingle, *Citropsis* (Engl.) Swing. & M. Kell., and *Murraya* Koenig ex Linn. (Esen and Soost, 1972; Soost and Cameron, 1975; Cameron and Soost, 1984). However, polyploids are known to occur. Many spontaneous and induced tetraploids in *Citrus* and *Poncirus* have been reported, and a few triploids have also been identified in *Citrus* (Cameron and Frost, 1968; Soost and Cameron, 1975). More than ten different types of aneuploids have been reported (Esen and Soost, 1972).

The size of the *Citrus* genome is small in comparison with many other plant species (Price, 1976; Levin and Funderburg, 1979; Guerra, 1984; Ohri and Kumar, 1986). Based on the study of *Citrus sinensis*, the citrus genome size (C value) was determined to be 0.6 picogram (pg) per haploid DNA content (Guerra, 1984), or equal to 5.63 x 10^5 kb of DNA. Thus, the *Citrus* genome is the same size as *Oryza*

sativa (0.6 pg/C), similar to *Lycopersicon esculentum* (0.7 pg/C), and much larger (ca. 9x) than *Arabidopsis thaliana* (0.07 pg/C); however it is smaller (ca. 11% as large) than *Zea mays* (5.5 pg/C) (Bonierbale et al., 1988; McCouch et al., 1988; Zamir and Tanksley, 1988).

Mutations in Citrus. Spontaneous mutations occur frequently in *Citrus* (Soost and Cameron, 1975). The most noticable mutations are bud mutations that result in detectable limb sports (Cameron and Frost, 1968; Cameron and Soost, 1984). Somatic mutation followed by the persistence of both the old and new cell types commonly produces chimeras, which have long been recognized in *Citrus* (Soost and Cameron, 1975). A recent study carried out in Florida showed that the frequencies of fruit chimera occurrence in several varieties ranged from 0.041% to 0.271% (Bowman et al., 1989).

From a long series of studies on bud variation, Shamel (1943) reported that the frequencies of entire-tree variations of 'Washington Navel' and 'Valencia' oranges and 'Eureka' lemons ranged from less than 10 per cent to more than 75 percent, depending upon the orchard. Interestingly, Shamel also reported that limb variants of the same types were often found repeatedly, suggesting that the same mutations were sometimes re-occurring. The variant types of several different varieties of orange, lemon and grapefruit were also studied. The kinds of changes observed were found to be similar among all varieties.

Most mutations are unfavorable, but some highly valuable mutations have occured. New varieties selected from bud mutations have been reported for many different cultivated *Citrus* groups (Soost and Cameron, 1975). Mutations resulting in the rise of new varieties are famous in some *Citrus* groups. One example is the 'Washington' navel orange, a sport from the 'Selecta' orange, which is noted for its tendency to give rise to bud variations. There are numerous varieties derived from

'Washington' navel during the past century (Hodgson, 1967; Cameron and Frost, 1968).

Satsuma mandarin (*C. unshiu* Marc.) is another *Citrus* group in which bud variation has been very important for creating new varieties (Hodgson, 1967). Many Satsuma mutations have been found and studied in Japan since the early 20th century. One special type of mutation in Satsuma called Wase was identified from three different varieties. One Wase Satsuma, called Kuwano Wase, exhibits frequent reversion to the Owari-type Satsuma from which it was derived. The mechanism involved in Satsuma bud variations is unclear. Although different hypotheses have been suggested, such as single-gene mutations or the shifting in histogenic layer composition of a chimera, there is not agreement among scientists (Cameron and Frost, 1968).

Inherited characters. Little is known about inherited characters in *Citrus*. This may be partly because most *Citrus* species are highly heterozygous. The F1 progeny often display a wide quantitative range of character expression, but few traits have been well studied. Some documented quantitative characters include cold hardiness, inheritance of leaf and rind oils, tolerance to root rot caused by *Phytophthora* spp., and chloride resistance (Soost and Cameron, 1975). Fruit acidity has been found to be inherited in a semiquantitative manner in some crosses (Cameron and Frost, 1968). Tolerance of citrus nematode (*Tylenchulus semipenetrans*) and burrowing nematode (*Radopholus similis*) are both heritable (Soost and Cameron, 1975). Resistance to California red scale was shown to be related to parentage. The dwarf trait shows discrete segregation, but its inheritance pattern is still not clear (Soost and Cameron, 1975).

A few characters appear to be under the control of one or a few genes. Nucellar embryony, common to most *Citrus* species, is postulated to be controlled

by one dominant gene; this conclusion is based on crosses of monoembryonic and polyembryonic parents of several varieties. Monoembryonic varieties are homozygous for the recessive allele (Cameron and Frost, 1968). The trifoliate leaf character of *Poncirus* shows complete dominance over the monofoliate leaf of *Citrus*. Although its segregating pattern does not suggest single gene control, it is most likely there are two principle genes involved in this trait (Soost and Cameron 1975). Toxopeus (1962) reported another gene system controlling the trifoliate leaf character within the genus *Citrus*, other than from *Poncirus*. He studied the cross of *C. grandis* (L.) Osbeck x *C. hystix* DC (both monofoliate) and found one third of the progeny were trifoliate. This phenomenon led him to suggest that two complementary dominant genes may be involved, with each species being heterozygous for one of them. In the same paper, using the cross *C. grandis*. X 'Meyer' lemon (*Citrus* hybrid), he reported that the purple color characterictic of young lemon growth may be controlled by a dominant gene. Using ELISA (Enzyme-Linked Immunosorbent Assay) techniques and suitable experimental populations derived from crosses between *Poncirus trifoliata* and other species, the citrus tristeza virus (CTV) resistance trait was found to be dominant and assumed to be controlled by a single gene (Yoshida, 1985). Polyphenol oxidase-catalyzed browning of young shoot extracts was studied as an indicator in *Citrus* taxonomic studies and for identification of zygotics among nucellar seedlings (Esen and Scora, 1975; Esen et al., 1975). Browning was found to be dominant to nonbrowning. The presence of substrate is under single locus control (Esen & Soost, 1974a, 1977; Torres et al., 1978).

Biochemical and molecular studies. Many biochemical markers have been studied for use in the identification of zygotic and nucellar seedlings following hybridization with polyembryonic seed parents. These methods include infrared

spectroscopy of leaf oil (Pieringer and Edwards, 1964), assays of polyphenol oxidase-catalyzed browning of young shoot extracts (Esen and Soost, 1974a, 1974b), total protein analysis by SDS-PAGE, and non-histone protein analysis by PAGE (Ortiz et al. 1981). Ortiz et al. (1981) concluded that the composition of biochemical compounds was less influenced by environmental factors than were morphological characters. The biochemical tests can usually be performed in an early stage of plant development. Unfortunately, with the exception of the polyphenol oxidase-catalyzed browning method, most studies of these biochemical traits were unable to provide a clear genetic understanding about the traits, thus limiting their further application.

The most useful biochemical approach is the use of enzymatic markers, i.e. isozyme analysis. At least ten *Citrus* isozymes have been characterized genetically and have been successfully applied to genetic, taxonomic, and breeding studies (Iglesias et al., 1974; Soost and Williams, 1980; Cabera and Lima, 1981; Hirai and Kozaki, 1981; Soost and Torres, 1981; Torres et al., 1982). Leaf isozymes have provided the largest category of genetic markers thus far in *Citrus*, and are most suitable for gene mapping. Torres (1983) listed eight isozyme genotype systems specified by ten genes with 35 alleles within *P. trifoliata* and eight cultivated *Citrus* species. These isozyme systems include glutamate oxaloacetate transaminase (GOT-1, 4 alleles; GOT-2, 3 alleles), phosphoglucose isomerase (PGI-1, 3 alleles), phosphoglucose mutase (PGM, 5 alleles), malate dehydrogenase (MDH-1, 2 alleles; MDH-2, 2 alleles), leucine aminopeptidase (LAP, 3 alleles), hexokinase (HK, 4 alleles), isocitrate dehydrogenase (IDH, 4 alleles), and malic enzyme (ME-1, 5 alleles). Roose (1988) identified more useful enzyme systems that have been studied in *Citrus*, including esterase (EST, 1 locus), acid phosphatase (ACP, 1 locus), superoxide dismutase (SOD, 1 locus), IDH (2 loci), Me(2 loci), and PGM (2 loci),

There are five enzyme systems with locus number only inferred from phenotypes, and therefore needing further confirmation by genetic analysis in the future. They are ribulose-1,5-biphosphate carboxylase (LSU, 1 locus), ribulose-1,5-biphosphate oxygenase (SSU, 1 locus), shikimate dehydrogenase (SDH, 1 locus), 6-phosphogluconate dehydrogenase (6-PGD, 2 loci), and GOT (3 loci). Isozyme analysis has also been used to confirm somatic hybridity of regenerants from protoplast fusion experiments with *Citrus* and related genera (Grosser et al., 1988a, 1988b).

Molecular study of *Citrus* at the DNA level has been minimal. However, some preliminary RFLP work has been started, and its potential application to problems in *Citrus* breeding and systematics has also been discussed (Roose, 1988). RFLP analysis through DNA-RNA hybridization also was employed to examine somatic hybrids between *Citrus sinensis* and *Poncirus trifoliata* (Ohgawara et al., 1985).

The first linkage study in *Citrus* was performed using nine enzyme systems specified by 12 loci, and nine families of *Citrus grandis* x *C. jambhiri* and *C. grandis* x *Poncirus trifoliata* (Torres et al., 1985). Two linkage groups were defined; one included two loci (*got-1* and *mdh-1*), and the other included three loci (*mdh-2, me-2* and *me-1*, in order). There have been no RFLP marker linkage studies in *Citrus* nor any study reporting linkage of biochemical or molecular markers to morphological or whole plant traits.

Conclusion

Genetic studies and variety improvement of tree crops have been hindered in the past by technical difficulties such as great heterozygosity, pollen and/or ovule sterility, cross/self-incompatibility, and the inability to create inbred lines; and economic factors including large individual plant size and long periods of juvenility (Adams, 1983; Galletta, 1983; Sherman and Lyrene, 1983). *Citrus* is one of the tree

crops for which classical genetic studies and conventional breeding efforts are especially difficult to perform because of the reasons outlined above, such as polyembrony and the limited information about the genetics of economically important traits. Recently, molecular biological technology has provided a great opportunity for overcoming the limitations of genetic studies of trees.

A brief review of the literature that focused on linkage mapping, the aspects of technologies that are available for molecular study and linkage mapping, and the state of *Citrus* genetics was made in this chapter. The purpose of the review was to build a connection between this old crop and the new technology. The research performed for this dissertation was an attempt to initiate molecular biological studies of *Citrus* through genome characterization and linkage map construction by RFLP and isozyme analyses, as described in the following chapters.

CHAPTER 3

GENOMIC LIBRARY CONSTRUCTION AND CLONE SCREENING

Introduction

Restriction fragment length polymorphism (RFLP) analysis is a powerful technique for plant genetic and breeding studies (Burr et al., 1983; Helentjaris et al., 1985, 1986; Roose, 1988; Tanksley et al., 1989). RFLP analysis requires an appropriate library from which clones of single or low copy number sequences can be selected for use as probes. A library that contains sequences, randomly interspersed throughout the nuclear genome, that frequently detect polymorphism will be an ideal source of probes. In many cases, e.g. when doing chromosome mapping, the exclusion of organellar clones is desirable. Although cDNA libraries and genomic libraries have been constructed as probe sources for some major crop plants, e.g. maize (Bernatzky and Tanksley, 1986; Helentjaris, 1987), tomato (Helentjaris et al., 1989), potato (Tanksley et al., 1989), lettuce (Landry et al., 1987), and rice (McCouch et al., 1988), this has yet to be accomplished for many other plant species. Techniques are needed to reduce the effort and expense of initial RFLP studies in new species.

Some approaches have been found to be helpful in achieving the objectives described above. Although cDNA libraries have been superior sources of useful probes for a few plant species, a genomic library used as a probe source should better represent the whole genome (Landry and Michelmore, 1985). Cloning under-methylated genomic sequences that are completely digested by a methylation-

sensitive enzyme, such as PstI, may help enrich for clones of single copy sequences in a genomic library (Helentjaris, 1987; Burr et al., 1988). Extraction of DNA from isolated nuclei, or screening clones derived from total DNA preparations by hybridization with labeled organellar DNA, can eliminate clones showing uniparental inheritance (Landry and Michelmore, 1985; Burr et al., 1988). Short DNA inserts cloned in plasmid vectors may have less chance of containing interspersed repeat sequences than inserts cloned in lambda or cosmid libraries (Landry and Michelmore, 1985). Finally, the dot blot hybridization technique described by Landry and Michelmore (1985) may help distinguish clones of repetitive sequences.

This chapter describes protocols developed in the course of defining an RFLP map for *Citrus*, a genus for which little genetic information exists. A simplified procedure for creating a genomic library that is low in organellar clones and enriched for single copy genes is detailed. Also described are procedures that make the selection of appropriate probes for plant RFLP studies more efficient.

Materials and Methods

Plant Materials

Plant material for making the genomic library. Young leaves of 'Temple' tangor [*C. reticulata* Blanco hybrid] were used to isolate genomic DNA for constructing a *Citrus* genomic library. Healthy, newly expanding leaves, <3 cm long, were collected from the field, shipped in ice to the laboratory, rinsed with deionized water, and immediately used for procedure 1 (see Experimental Procedures in this chapter).

Plant materials used as screening filters in library survey. Three *Citrus* lines were used for the clone copy number survey through Southern analysis in this study. They were: 1) *Citrus reticulata* cv. Clementine (called 'Clementine'

hereafter in this dissertation); 2) *C. paradisi* Macf. cv. Duncan (called 'Duncan' hereafter); and 3) a hybrid of these two varieties called LB1-21, which was derived from 'Clementine' x 'Duncan'. These materials were obtained from Fred G. Gmitter at the Citrus Research and Education Center (CREC), University of Florida. They were all healthy adult trees growing in the field under normal conditions. No stresses could be recognized when leaf samples were collected. Young leaves were collected from these three lines for DNA isolation.

Experimental Procedures

Several different protocols were employed in this study. Some were for library construction (Procedures 1-4), while others were used for the survey of the library (Procedures 5-7).

Procedure 1: DNA isolation for cloning.

1) Fresh plant leaves (30 g) were collected for DNA extraction. The leaves were cut into strips ca 0.5 cm wide.
2) The samples were treated with 50% ethanol for 3 min to stablize nuclear membranes, followed by 3 rinses with deionized water. For convenient handling, the samples were wrapped with cheesecloth. After this step, samples could be stored at -8°C for several weeks.
3) Five volumes (in this case, 150 ml) of ice cold (4°C) EB (Extraction buffer: 100 mM Tris-HCl, pH 8.0, 0.5 M sucrose) were added, and the sample was macerated in a blender for 60 sec. The mixture was rapidly poured into an appropriate plastic container for further maceration with a Polytron for 45 sec.

4) The homogenate was then quickly poured through a series of filters: 14- and 28- mesh sieves, 4 layers of cheesecloth, and 210μ, 105μ, and 45μ nylon screens.
5) The filtrate was collected in two 40-ml Oak Ridge tubes and the volume in each tube was brought to 40 ml with ice cold EB. ß-mercaptoethanol (0.4 ml) was added to each tube.
6) The sample was pelleted at 300 x g, 4°C, for 5 min, resuspended in EB, and 1/50 volume of 25% Triton X-100 was added, followed by gentle mixing.
7) The mixture was pelleted at 500 x g, 4°C, for 15 min. The pellet was expected to be white instead of dark green, indicating lysis of chloroplasts and pelleting of nuclei. The pellet was resuspended in 3 ml Resuspension buffer (RB: 100 mM Tris-HCl, pH 8.0, 20 mM EDTA-Na; use 1 ml buffer per 10 g leaves).
8) An equal volume (3 ml) of Lysis buffer (100 mM Tris-HCl, pH 8.0, 2% SDS) was added and the mixture was swirled gently to lyse nuclei.
9) Next 3 ml of phenol and 3 ml of chlroform/isoamyl alcohol (24:1), were added, followed by gentle mixing and centrifugation at 2000 x g, 25°C, 15 min.
10) The aqueous phase was removed to two cellulose nitrate tubes. To each ml of sample, 1 gm of solid cesium chloride and 80 ul of ethidium bromide stock (10mg/ml) were added. The samples were mixed gently and centrifuged at 171,000 x g (Beckman 70.1 Ti Rotor, 50,000rpm), 23°C for 25 hr.
11) The DNA band was extracted and the cesium gradient purification was repeated as in the previous step.

12) The DNA band was collected, ethidium bromide was removed, and the sample was dialyzed vs TE as described by Maniatis et al (1982). The DNA was quantified spectrophotometrically and stored at -20°C.

Procedure 2: Selection of DNA fragments for ligation.

1) After dialysis, without further purification, the DNA was restricted with *Pst*I. In this study, 250 μg DNA and 120 units of enzyme were used for restriction.
2) The DNA was purified by phenol and phenol/chloroform extraction, followed by ammonium acetate (8M, 1/10 volume) and ethanol (100%, 2 volumes) precipitation. Rehydration of the DNA was accomplished in 0.1 x NTE (1 x NTE: 10 mM NaCl, 1 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, pH 8.0).
3) DNA fragments were sized by centrifugation in a continuous sucrose gradient (10-30% w/v sucrose in 1 x NTE, 78,000 x g, 15°C, 20 hr, use SW41 Rotor or equivalent). The solution was divided into 40 fractions and 25 μl samples were taken from each fraction to check DNA fragment size by gel electrophoresis (0.8% agarose). The fractions containing DNA of 500-2,500 base pairs (bp) were chosen for ligation.

Procedure 3: Making recombinant plasmid. This study used pTZ18R, produced by Pharmacia LKB Biotechnology, Inc., as the DNA cloning vector. This vector is capable of serving as a phagemid, but was used only as a plasmid vector throughout this study. The plasmid contains a short DNA segment coding for the α-peptide of β-galactosidase, an ampicillin resistance gene, and a polylinker sequence similar to that of the pUC18 plasmid.

The plasmid was restricted with *Pst*I and then dephosphorylated with alkaline phosphatase in order to remove the 5' terminal phosphate, thus avoiding self-

ligation and enhancing the cloning efficiency (Perbal, 1984). The linear and dephosphorylated plasmid was then ligated to *Citrus* insert DNA, prepared by procedure 2, using T4 ligase (produced by Bethesda Research Laboratories). The ligation procedure was according to Perbal (1984).

Procedure 4: Transforming the recombinant plasmid into *E. Coli* TB1. The procedure of transformation includes two main steps, the preparation of competent bacterial cells and the transformation of plasmids into the cells. This study used *E. coli* TB1 strain as the host bacterium, which is genotypically characterized as being lacZ$^-$, amp$^-$ and plasmid free. The preparation of competent bacterial cells and the transformation process were according to Hanahan (1983). LB recipe containing ampicillin antibiotic was used during the procedures. X-gal (5-chloro-4-bromo-3-indolyl-β-D-galactoside) and IPTG (isopropylthio-β-galactoside) screening methods were used for the identification of desired transformants (Vieira and Messing, 1982).

Four different controls (treatments 1-4 in Table 3-1) were designed to monitor the experimental process. The results were recorded after overnight incubation (37°C) of the transformed cells. Under the X-gal-IPTG-Amp screening filter, only white colonies were desired. Non-transformed cells would be killed with the previously added antibiotic, i.e. ampicillin; transformed cells harboring no insert would grow into blue colonies due to the reaction between X-gal and the still functional β-galactosidase.

Procedure 5: Rapid clone screening. To enhance the rate and efficiency of clone screening, clones were screened using the alkaline lysis method described by Maniatis et al. (1982), but the procedure was divided into two sections:

1) Recombinant (white) colonies from procedure 4 were selected for screening. Plasmid DNA was extracted via the alkaline lysis method but the procedure was stopped after the phenol/chloroform extraction step.

For convenience, this was called the Short Miniprep (SMP). The mixture produced at this point in the procedure could be stored in the extraction tube at -20°C for several weeks after microfuging.

2) The approximate sizes of the plasmid inserts were determined by running 5 μl of supernatant from each sample on a 1.5% agarose gel. Plasmid that did not contain insert was extracted by SMP from an extremely blue colony and was used as a size marker in the gel. Plasmids that migrated significantly more slowly than the marker plasmid and thus putatively contained large inserts were selected. These samples were put through an additional phenol/chloroform extraction and the remaining steps of the alkaline lysis method.

3) The plasmids from selected clones were restricted with *Pst*I and the DNA was separated in an 0.8% agarose gel. Clones having inserts larger than 500 bp were named as pgCit followed by a 3-digit number (i.e. pgCit001 and so on; if there was a double insert, the large clone was denoted with 'L' and small clone with 'S' after the number) and subsequently used as probes for Southern analysis.

<u>Procedure 6: DNA isolation for Southern blotting</u>. To isolate DNA from parent trees and progeny populations for Southern blotting and RFLP analysis, the procedure of Dellaporta et al (1983) was used with the following modifications:

1) A larger amount (2-4 g) of leaf tissue was used as starting material.
2) The supernatant was poured through a Miracloth filter into a 30-ml tube and 10 ml isopropanol was added. The tube was incubated at 4°C overnight. This reduces the precipitation of polysaccharides.

3) The DNA was redissolved with 50 mM Tris and 10 mM EDTA. Proteins were removed by extracting once with phenol, once with a 1:1 mixture of phenol and chloroform, and once with chloroform.
4) The DNA was precipitated with 0.5 volume of 7.5 M ammonium acetate and 2 volumes of absolute ethanol.

Procedure 7: Restriction digests, electrophoresis, and Southern analysis. Southern analysis includes: 1) the blotting procedure that transfers the electrophoretically separated DNA onto a membrane (blot) to which the probe DNA is capable of hybridizing; and 2) detecting hybridization by the signal of a labeled probe.

For making blots, six 6-bp-recognition restriction enzymes (*Eco*RI, *Eco*Rv, *Bgl*II, *Hin*dIII, *Bam*HI and *Pst*I) were employed to digest the *Citrus* DNA prepared by Procedure 6. Using 2 μg DNA per lane, the DNA digests were separated in 1% agarose gels according to Maniatis et al (1982). The gel was 20cm(w) x 25cm(l) x 6mm in dimension, and was run at 60 volts for 16-18 hr before it was subjected to blotting. Blotting procedures were also according to Maniatis et al (1982) and used 10 x SSC solution as blotting buffer. Hybond-N nylon sheet was used as transfer membrane (blot). The blot was illuminated with 254 nm UV light for 4 min right after the blotting process to induce covalent DNA-blot binding.

For making DNA probes, insert DNAs (selected by Procedure 5) were separated in 0.8% agarose gels and purified by GENECLEAN (BIO 101, Inc.). A Random Primer DNA Labeling kit (Boehringer Mannheim) and P^{32} CTP (Deoxycytidine 5'-triphosphate, tetra [triethylammonium] salt, [α-^{32}p]) nucleotide were then used to label the probes. The prehybridization and hybridization procedures were according to Church and Gilbert (1984).

The method used as basic procedure for washing was two washes with 0.5 x SSC, 0.1% SDS, 65°C, 30 min; followed by another two washes with 0.1 x SSC, 0.1% SDS, 65°C, 30 min. Such stringency is capable of determining about 99% homology of DNA sequences (Beltz et al., 1983; McCouch et al., 1988). After washing, the blot was exposed to X-ray film with 1 or 2 intensifying screens, and kept at -80°C for 1-7 days depending on the hybridization signal.

Results and Discussion

DNA Preparation and Genomic Library Construction

DNA isolated for cloning. Under the conditions described in Procedure 1, the concentration of 'Temple' DNA prepared for cloning was 125 ng/μl, determined spectrophotometrically, and the yield of DNA was 13.9 μg per g fresh weight. Spectrophotometric determination showed that the ratio between OD readings at 260 nm and 280 nm was 1.7785, indicating extremely high purity (compared to 1.8 for pure DNA, Maniatis et al., 1982). The purity and concentration of the DNA was suitable for enzyme restriction without further precipitation and purification, and the restriction digest by *Pst*I was suitable for ligation after sucrose gradient centrifugation.

In preliminary experiments, *Citrus* DNA samples derived from leaf tissue gave poor OD260/OD280 readings (for example 1.3) with only one cesium gradient centrifugation. Also, ethanol precipitation following one cesium gradient centrifugation resulted in the loss of most of the DNA due to difficulty in rehydrating the DNA pellet. These observations suggest that some unknown cell constituents in *Citrus* may be very difficult to separate from DNA and so interfere with purification. Studies have shown that polysaccharides and tannins cause a major problem in plant DNA isolation as well as in cloning of plant DNA (Murray

and Thompson 1980). The success acheived with Procedure 1 suggests that for species where DNA isolation is especially difficult, two cesium gradient centrifugations may help solve the problem of DNA purification. DNA prepared with this protocol was at least 50 kb in size (Fig. 3-1), a length suitable for enzyme restriction and the creation of unbiased restriction fragments for genomic library construction.

The genomic library. Table 3-1 summarizes the overall results obtained in the library construction experiment. About 970 white colonies (in 10 plates) were identified and established for further tests. This number of colonies was sufficient for initiation of a gene mapping project.

Transformation efficiencies, expressed as number of colonies formed per picogram or microgram of plasmid DNA may differ due to *E. coli* strain, plasmid, insert size, and transformation procedure used (Hanahan, 1983; Old and Primrose, 1985). In this study, the concentration of pTZ18R plasmid in each plate was ca. 20ng. Based on the result of treatment 2 in Table 3-1, the transformation efficiency can be calculated as 1.4×10^5 transformants per microgram plasmid DNA. This is significantly lower than the optimal efficiencies (between 10^7-10^8 transformants/μg plasmid) summarized from most *E. coli* K12 strains (Old and Primrose, 1985). Thus far there is no information about the optimal efficiency of the TB1/pTZ host-plasmid combination. The factor(s) that led to this low efficiency are not clear.

None of the control treatments (1-4, Table 3-1) contained insert DNA; therefore there should be no white colonies in these plates. However, in this experiment, 3.5% and 18.8% of colonies showed white color in treatments 3 and 4 respectively. One possible explanation for this is that some white colonies might be caused by sequence changes in the plasmid, e.g. deletion, during the

Fig. 3-1: DNA isolated with different procedures. Lanes 1, 2, and 3 are lambda DNA, *Citrus* DNA prepared with Procedure 1, and *Citrus* DNA prepared with the modified Dellaporta method (i.e. Procedure 6), respectively. Lane 4 is *Citrus* DNA prepared with Procedure 1 and restricted with *Eco*RI; and lane 5 is *Citrus* DNA prepared with Procedure 6 and restricted *Eco*RI. Lane 4 has a little more DNA than lane 5, but has only very faint distinct white bands in comparison with lane 5. The gel was run from top to the bottom in the picture.


1
2
3
4
5

Table 3-1: Results of transformation experiment for genomic library construction. Treatment 1 was used to detect contamination of bacterial cells. Treatment 2, 3, and 4 were to detect transformation, ligation, and dephosphorylation efficiency, respectively.

	Treat-ment	TB1 cells	Plasmid [1]			T4 ligase	Citrus insert	No. colonies per plate	Index	White colonies per plate	% white colonies	No. plates
			P1	P2	P3							
Controls	1	+[2]	–[2]	–	–	+	–	0	0	0	0	1
Controls	2	+	+	–	–	+	–	2852	100	0	0	2
Controls	3	+	–	+	–	+	–	1896	67	67	3.5	2
Controls	4	+	–	–	+	+	–	88	3	17	18.8	2
	5	+	–	–	+	+	+	180	6	97	53.8	10

[1] P1: uncut plasmid pTZ18R; P2: linear pTZ18R cut with PstI;
P3: linear pTZ18R cut with PstI and then dephosphorylated.
[2] + : added; -: not added.

ligation reaction. Normally T4 ligase is incapable of ligation when only the dephosphorylated linear plasmid is present, and therefore treatment 4 should not have yielded white colonies. Because gel electrophoresis indicated that the plasmid was completely restricted by *Pst*I enzyme (picture not shown), the result of treatment 4 may indicate that the dephosphorylation reaction in procedure 3 was not complete.

The white colonies in treatment 5 were transferred into microtiter trays that contained Freeze Broth medium in each well for storage. The Freeze Broth medium was prepared following Gergen et al. (1979). The trays were stored at -20°C.

Library Surveys

DNA prepared for Southern blotting. The yield of DNA obtained via the modified Dellaporta protocol from the plants to be used for the library survey ranged from 10 μg to 30 μg per g fresh weight of leaf tissue. The molecular weight of the DNA fragments was at least 50 kb (Fig. 3-1). In a few cases, extracted DNA was quite viscous, but all of the DNA samples were restrictable by the enzymes tested, and produced good band(s) in Southern analyses. When 2 μg of restriction digest per lane was used for Southern blotting, the blots could be reused up to 8 times.

Short miniprep enhances screening efficiency. Analyses using the alkaline lysis miniprep procedure (Maniatis et al., 1982) showed that only a small portion of clones in the library harbored inserts larger than 500 bp. These results were similar to those of some other researchers (Slightom and Drong, 1988). The reason for this was not clear. Perhaps it was because sucrose gradient centrifugation produced fractions without strict boundaries, i.e. there were quite a few smaller

fragments in the selected fractions of specific fragment sizes, and the bacterium more easily takes up a smaller plasmid.

For efficient screening and selection of clones with inserts of the desired size, colonies were subjected to the SMP protocol followed by gel electrophoresis (Fig 3-2), and the rate of migration of supercoiled plasmid DNA was compared to that of the size marker. Migration of a supercoiled plasmid in an agarose gel can faithfully reflect its molecular weight, especially in the absence of ethidium bromide (Maniatis et al., 1982). Therefore it also reflects the size of the insert if there is only one insert within the plasmid. It was found that the plasmid always migrated more slowly than the slowest migrating RNA (Fig. 3-2), making the gel easy to read and treatment with RNase unnecessary. Using SMP, it was possible to identify clones that potentially had large inserts and to estimate insert sizes without proceeding through the entire miniprep procedure. After screening the library with SMP, it appeared that about 32.2% of the library contained clones with inserts >500 bp. Further analysis revealed that about 82% of the promising colonies selected by this approach (or 26.4% of the whole library, Fig. 3-3) actually harbored inserts larger than 500 bp. The SMP protocol was quite economical in time and cost. Thus, the SMP is not only an approach for enhancing the efficiency of clone screening, but is also a quick procedure for the overall survey of the library immediately after its construction.

Evidence of low organellar DNA content in the cloned DNA. In higher plants, the chloroplast DNA in leaf cells is between 10 and 20% of the total DNA (Grierson and Covey, 1984; Scott and Possingham, 1982). A higher percentage of mitochondrial DNA may be present, because mitochondria are usually found in larger numbers than chloroplasts inside the same cell, and the plant mitochondrial genome is usually larger than that of chloroplasts (Wolfe, 1983; Grierson and Covey,

Fig. 3-2: The relation between the migrating speed of circular plasmid prepared by SMP and its insert size shows that SMP is a shortcut of the miniprep procedure for selecting large clones from the library. Photo A: Circular recombinant plasmid prepared by SMP, then run in a 1.5% agarose gel. The white smears in the lower part of lanes 2-8 are RNA. Photo B: Linear pTZ18R plasmid and insert excised by *Pst*I, then run in a 1.5% agarose gel. The migration is from top to the bottom in both 2 photos. Lane 0: 1 kb DNA ladder; lane 1: pTZ18R plasmid (from Pharmacia; M.W. = 2.9 kb); lane 2: plasmid from a very blue colony (without harboring insert) that serves as size marker in SMP screening procedure; lane 3: plasmid, containing a very small insert of ca. 200 bp, migrats slightly slower than the size marker (lane 2) and forms a very faint band in photo B); lane 4: a plasmid with 2 small inserts (each insert is < 0.4 kb); lanes 5, 6, 7, & 8: plasmids that contain inserts of ca. 0.5, 1.6, 2.5, & 4 kb, respectively. The numbers on the left hand side are the molecular weights in kb.

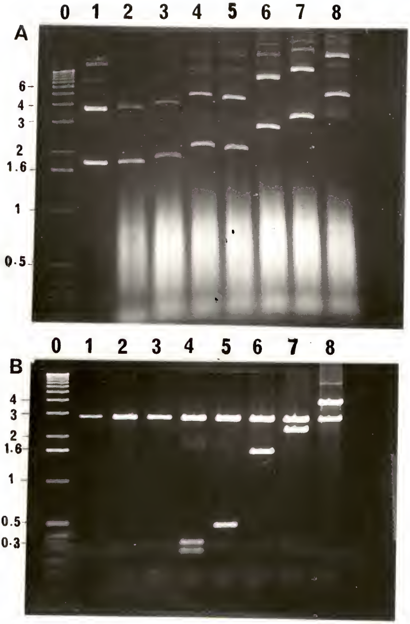
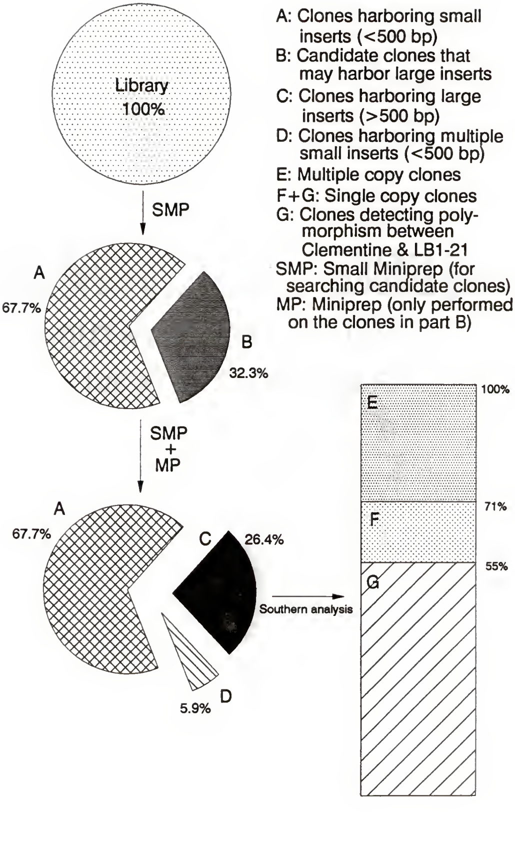

0 1 2 3 4 5 6 7 8
A
6
4
3
2
1.6
1
0.5
0 1 2 3 4 5 6 7 8
B
4
3
2
1.6
1
0.5
0.3

Fig. 3-3: Clone screening flow chart.


A: Clones harboring small inserts (<500 bp)
B: Candidate clones that may harbor large inserts
C: Clones harboring large inserts (>500 bp)
D: Clones harboring multiple small inserts (<500 bp)
E: Multiple copy clones
F+G: Single copy clones
G: Clones detecting polymorphism between Clementine & LB1-21
SMP: Small Miniprep (for searching candidate clones)
MP: Miniprep (only performed on the clones in part B)
Library
100%
SMP
A
67.7%
B
32.3%
SMP
+
MP
A
67.7%
C
26.4%
D
5.9%
Southern analysis
E
F
G
100%
71%
55%

1984). This makes the elimination of organelle DNA very important when constructing a genomic library. A straightfoward solution to this problem is nuclei isolation prior to genomic library construction, but the DNA yield is usually very low with this approach (Murray and Thompson, 1980). This was also found in preliminary experiments in this study. Alternatively, it is possible to screen clones derived from total cell DNA by hybridization with labeled organelle DNA (Burr et al., 1988). However, this is a time-consuming and labor intensive process. The results obtained in this study demonstrate that other strategies may be used for the elimination of organelle DNA.

In Procedure 1, the pellet was white after lysis with Triton X-100, in contrast to the green pellet obtained without the addition of Triton X-100 (data not shown). This indicated lysis of chloroplasts, and for the reasons described below, it was probable that all kinds of organelles were lysed. Fig. 3-1 shows that DNA prepared by Procedure 1 and restricted with *Eco*RI had fewer distinct bands (only 2 very faint bands, if viewed very carefully), in comparison with an analogous sample that unquestionably contains both nuclear and organelle DNA prepared with the modified Dellaporta protocol (Procedure 6). This result implied that at least most organellar contamination had been eliminated. Triton X-100 previously has been used to solubilize the plasma membrane in plant cells, although this mild and nonionic detergent can also be used to disrupt the nuclear membrane (Keim, 1987). However, since nuclear membranes have been shown to be less susceptible to disruption (Kislev and Rubenstein, 1980), the concentration of Triton X-100 is critical to the lysis of organelles and nuclei retention. One other possible explanation for preferential lysis of organelles versus nuclei in this procedure is that the 50% ethanol pretreatment was more effective in wrinkling nuclear membranes as opposed to organelle membranes, and thus nuclei remained intact in the presence

of a proper concentration of a mild detergent such as Triton X-100 (Vallejos, personal communication). When the serial filtration described in Procedure 1 was performed, microscopic examination showed that only leaf pieces comprising less than ca 20 cells were found in the final filtrate. Only samples filtered in this manner gave homogeneous white pellets after Triton X-100 lysis. When the final filtrate contained quite a few larger leaf pieces, comprising 40 or more cells, then it produced a variegated white-green pellet after Triton and before SDS lysis. This was perhaps because Triton does not penetrate to the cytoplasm of cells in the central part of the larger leaf pieces as does the stronger detergent SDS. This indicated that filtration was also a crucial step in this protocol.

A study through Southern analyses indicated that 71% (41/58) of the clones selected from the genomic library were most likely single copy sequences, based on the hybridization banding pattern. Furthermore, in a survey of three genetically close *Citrus* types, i.e. 'Clementine', 'Duncan' and their hybrid LB1-21, 55% (32/58) of the clones were polymorphic and appeared to follow Mendelian inheritance, and were therefore concluded to have derived from the nuclear genome. Banding patterns indicated that 3% of the clones may have been maternally derived (i.e. organellar in origin). The remaining clones revealed no polymorphism. In another survey of species from different, more divergent genera (*C. grandis* [L.] Osb., *Poncirus trifoliata* [L.] Raf., and a hybrid of these two species), 76% (13/17) of the tested clones were found to be polymorphic and inherited in a Mendelian fashion (R. Durham, personal communication). These results reflect good elimination of organelle DNA.

Overall Features of the Constructed Genomic Library

Using a series of strategies designed to reduce organelle clones and to increase single copy clones, a *Citrus* genomic library comprising ca. 970 clones was constructed and preliminarily surveyed in this study. The time-saving procedure developed in this study, i.e. Short Miniprep, revealed that about one third (32.2%, Fig 3-3) of the colonies potentially harbored large inserts (>500 bp). However, a further survey by the miniprep procedure showed that only about one quarter (26.4%, Fig. 3-3) of the colonies harbored inserts larger than 500 bp and were good candidates for further studies. Southern analysis indicated that 18.7% of the library (i.e. 71% of the large inserts, ca. 180 clones) were most likely single copy sequences. Southern analysis also implied that about 14.5% of the colonies (i.e. 55% of the large inserts, ca. 140 clones) could be expected to serve as polymorphic clones when 'Clementine' mandarin and the hybrid LB1-21 were used on screening filters.

CHAPTER 4

CITRUS GENOME CHARACTERIZATION BY RFLP ANALYSIS

Introduction

Although many DNA clones were created (Chapter 3), only those clones with single or low copy numbers in the genome would be useful for genetic studies, especially linkage analysis and mapping. Features such as clone copy numbers largely depend on the characteristics of the genome being studied (Klug and Cummings, 1983) and, in turn, a survey for copy number profile and RFLP using random clones should reveal genome characteristics. Such information would be useful for further understanding of *Citrus* genetics. This chapter describes experiments designed to characterize the clones that detected polymorphism and to study the copy number profile of the clones. However, the *Pst*I-derived genomic library might not well represent the whole *Citrus* genome, because DNA fragments used for cloning were generated by a methylation-sensitive enzyme that would have selectively restricted the DNA. Therefore, an experiment was performed to determine the extent of methylation of the *Citrus* genome. Some RFLP features were found to be worthy of special notice. This chapter includes detailed discussion of these features and proposes models to interpret the noteworthy characteristics. The selected clones that detected RFLP between 'Clementine' mandarin and the hybrid LB1-21 were used as genetic markers for the progeny analysis and linkage study described in Chapter 5.

Materials and Methods

Plant Materials

The plant materials used as the screening filter ('Clementine', 'Duncan', and the F1 hybrid LB1-21) and the method used for DNA isolation for Southern analysis were the same as described in Chapter 3.

Determination of DNA Methylation

For determining the extent of DNA methylation in the *Citrus* genome, 400 ng of 'Duncan' DNA (isolated by Procedure 6 described in Chapter 3) was digested with *Pst*I and *Eco*RI, respectively. Fragments were separated in a 1.5% agarose gel, and stained with ethidium bromide to compare the difference in restriction. To determine the degree of digestion, 150 ng of circular pTZ18R plasmid DNA (which has a unique restriction site for each of these two enzymes) was added to each tube before restriction.

Clones Used for RFLP Survey

Seventy-nine clones ranging in size from 0.5 to 5 kb were selected from the genomic library described in Chapter 3 for use in the RFLP survey. Twenty cDNA clones, including 18 obtained from Dr. M. Roose (University of California at Riverside) and 2 from Mr. R. Durham (University of Florida), were also used for the RFLP survey. The clones from M. Roose were derived from Rough lemon (*Citrus jambhiri*) cDNA, cloned into pUC9 plasmid and transformed into *E. coli* JM83. They were named as pRLc followed by a 2-digit number. The clones from R. Durham were derived from *Poncirus trifoliata* cDNA, cloned into pTZ18R plasmid and transformed into *E. coli* TB1. They were named as pcPt followed by a 3-digit number.

The probe DNA prepared from *E. coli* clones and the procedures used for restriction digests, agarose gel electrophoresis, and Southern analysis were the same as described in Chapter 3. Six restriction endonucleases with different 6 base pair recognition sites (*Eco*RI, *Bgl*II, *Eco*RV, *Bam*HI, *Hin*dIII, and *Pst*I) were used for restriction digests of DNA from the two parents and the hybrid to make a blot screen. In some cases, only *Eco*RI, *Bgl*II, and *Eco*RV were used.

Modification of the Washing Method Used for Southern Analysis and RFLP Determination

The technique of Southern analysis and the basic washing protocol were the same as described in Chapter 3. In some cases, a modification of the washing method was utilized with clones showing faint hybridization besides the clear bands (a trace of extra loci) that could not be confirmed with high stringency wash conditions (0.1 x SSC, 0.1% SDS, 65°C). A wash of moderate stringency (using 1 x SSC, 0.1% SDS, 65°C for the last two washes) was done using another probing reaction with these clones to determine whether the faintly hybridizing bands could be better visualized under these conditions. In some other instances blots were first washed with moderate stringency conditions and exposed to X-ray films with 2 intensifier screens per film. After the films had been developed (usually 24-36 hr), blots were washed again with high stringency conditions, and re-exposed to compare the gene copy numbers and polymorphisms revealed by different stringency conditions. Also, in some instances, a rather high stringency (using 0.05 x SSC, 0.1% SDS, 65°C for the last two washes) was utilized to determine the stringency tolerance of *Citrus* DNA.

Results and Discussion

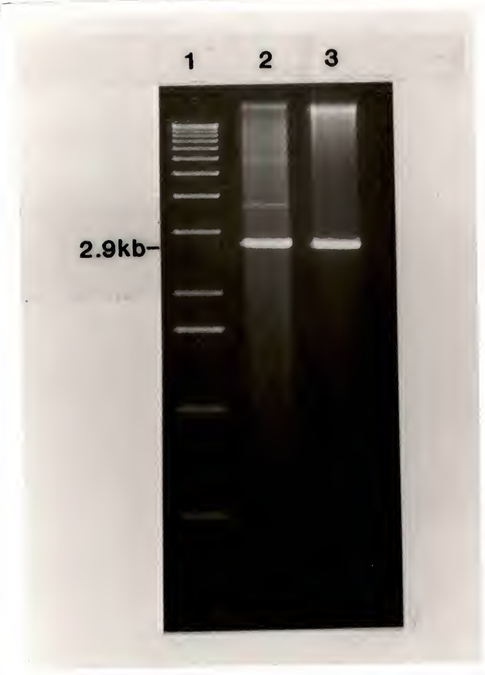
DNA Methylation in the Citrus Genome

*Pst*I recognizes the sequence 5'CTGCAG-3' and is sensitive to cytosine methylation in the 5' location; whereas EcoRI recognizes 5'GAATTC-3' and is methylation insensitive (McCouch et al., 1988). Fig. 4-1 shows that when *Citrus* DNA is digested with *Pst*I, a substantial portion remains as fragments of high molecular weight, but the distribution of fragments along the lane is more uniform following EcoRI digestion. The unique band of linear pTZ18R plasmid in each lane indicates the complete restriction with each enzyme. Since each lane had the same quantity of DNA, the result implies that *Pst*I recognizes fewer restriction sites in the *Citrus* genome than *Eco*RI. Thus it appears that the *Citrus* genome contains abundant methylated DNA. Such a result is consistent with observations where most methylation sensitive enzymes recognize few sites in maize DNA, because the maize genome contains a large amount of methylated DNA (Burr et al., 1988). Although the relationship between the repetitive sequences and the DNA methylation mechanism in the eukaryotic genome is not clear, it has been found in maize and tomato that fragments derived from *Pst*I digestion most likely are not from the the methylated areas in the genomes, and are rich in single copy sequence (Burr et al., 1988). Thus, it could be a crucial strategy to construct a *Citrus* genomic library using *Pst*I-derived DNA fragments, because a library rich in single copy clones is most preferrable in RFLP studies.

Clone Copy Number

Clones were classified into three categories, following McCouch et al. (1988) based on their copy numbers as revealed by Southern analysis. However, the method of classification and the category definition were modified as follows: (1)

Fig. 4-1: *Citrus* DNA fragments derived from *Eco*RI (lane 2) and *Pst*I(lane 3) restriction and separated in a 1.5% agarose gel. A substantial portion remains as fragment of high molecular weight in the *Pst*I digest, but the distribution of fragments along the lane is much more uniform in the *Eco*RI digest. The DNA was isolated from young leaves of 'Duncan' grapefruit using a modified Dellaporta method described in Chapter 3. The amount of 'Duncan' DNA in each lane (2 and 3) is 400 ng. Lane 1 is 1 kb DNA ladder used as size marker. The 2.9-kb bands in lane 2 and 3 are the linear plasmid pTZ18R that were added (as circular form) before enzyme reactions to indicate complete restrictions.


1
2
3
2.9kb

single copy sequences with no more than 4 clear bands per lane, that appeared to follow Mendelian inheritance patterns in 'Clementine' mandarin, 'Duncan' grapefruit, and their hybrid LB1-21; (2) multiple copy sequences with 3 to 10 bands present per lane; and (3) repetitive sequences with more than 10 bands per lane or with smeared signals with no discrete bands. Some examples of Southern analysis showing single copy, multiple copy, and repetitive sequences are shown in Figs. 4-2 and 4-3. The copy number classification of all the clones studied are listed in Appendix 1.

Effect of stringency conditions on copy numbers. A portion of eukaryote genes are known to be members of multigene families that often cross-hybridize in Southern analysis to produce multiple banding patterns because of sequence homologies (Beltz et al., 1983; Watson et al., 1987). Both closely and distantly related genes can be found within some multigene families. Typically the sequence corresponding to a central protein domain is substantially conserved within a family, but flanking sequences that encode protein "arms" (NH_2- and COOH-terminal domains) are much more variable (Beltz et al., 1983). In this study, the moderate stringency conditions used for the final wash in Southern analysis allowed detection of sequence homology between duplexes at a minimum level of 80%; the high stringency conditions used could detect only >99% homology (Beltz et al., 1983; McCouch et al., 1988). By controlling the hybridization stringency in Southern analysis, it was possible to reveal different banding patterns and different clone copy numbers for some clones. To study whether there are multigene families in the *Citrus* genome, 24 *Citrus* clones, including 15 genomic and nine cDNA clones, were assayed with both moderate and high stringency washes (see Appendix 1). Sixteen clones (11 genomic and five cDNA clones) revealed single copy numbers with the high stringency wash. With moderate stringency, two of these 16 clones (both were

Fig. 4-2: Southern analysis revealed single copy, multiple copy, and repetitive sequences in clones pRLc038, pgCit040, and pgCit031L, respectively. All of the blots were washed with the high stringency condition (0.1 x SSC, 65°C).

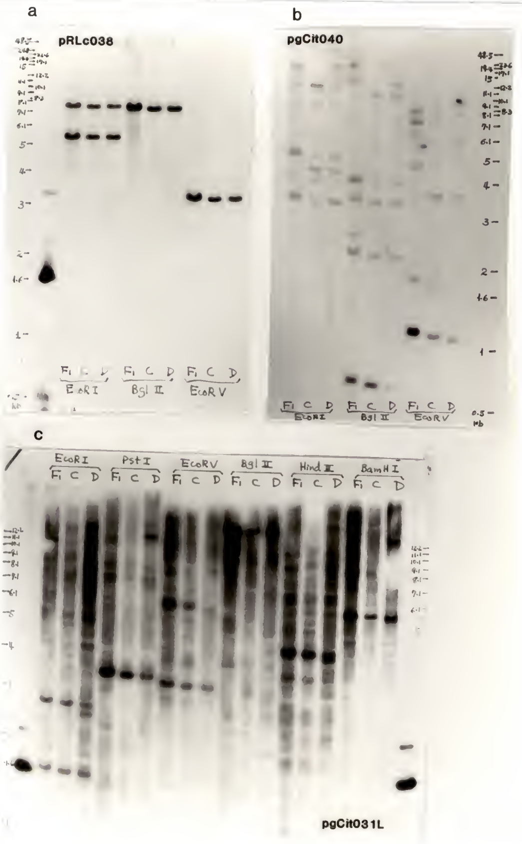
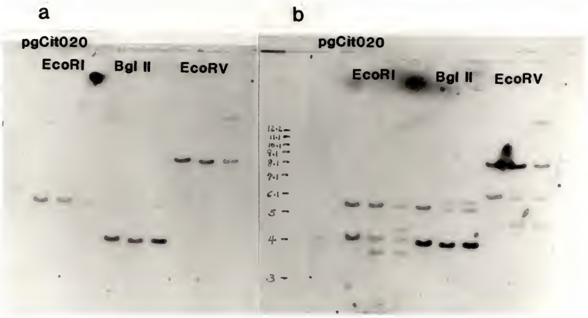

a
pRLc038
F1 C D
EcoRI
BglII
EcoRV
b
pgCit040
F1 C D
EcoRI
BglII
EcoRV
c
EcoRI
PstI
EcoRV
BglII
HindIII
BamHI
F1 C D
pgCit031L

Fig. 4-3: The genomic clone pgCit020 revealed a single copy sequence in the high stringency washing (a), but multiple copy sequences were found in the moderate stringency washing (b). Only the moderate stringency displayed polymorphism.


a
pgCit020
EcoRI
Bgl II
EcoRV
b
pgCit020
EcoRI
Bgl II
EcoRV
12.2
11.1
10.1
9.1
8.1
7.1
6.1
5
4
3

genomic clones) displayed multiple copies (one of them is shown in Fig. 4-3), one cDNA clone revealed repetitive sequences, and 13 (nine genomic and four cDNA clones) remained as single copy. Seven of these 24 clones (four genomic and three cDNA clones) showed multiple copy numbers in high stringency conditions; two of these seven (one genomic and one cDNA clone) revealed repetitive sequences with moderate stringency conditions. One of the 24 clones (cDNA) displayed repetitive sequences with both high and moderate stringency conditions. In total, five (three genomic and two cDNA clones) of the 24 studied clones (20.8%) revealed greater copy numbers under moderate, rather than high, stringency conditions. These results suggest that some sequences in the *Citrus* genome have evolved into multigene families.

After initial assays with high stringency washes, and following exposure to X-ray film, some blots were washed again with still higher stringency conditions (0.05 x SSC, 0.1% SDS, 65°C). The X-ray film exposures following these washes showed that 0.05 x SSC removed all of the probe DNAs from the blots, indicating that these conditions were too stringent for *Citrus* DNA. To confirm this conclusion, two blots were re-probed with two clones (one clone each blot) that showed clear bands with 0.1 x SSC washes and were then washed directly with 0.05 x SSC stringency. The X-ray film exposures yielded unreadably poor results.

Different species differ in their limits of stringency conditions. The stringency condition commonly used for the final washes in maize is 0.1 x SSC (Helentjaris et al., 1985; 1986), while 0.05 x SSC has given satisfactory results for rice and tomato (McCouch et al., 1988; Zamir and Tanksley, 1988). This study showed that the stringency tolerance of *Citrus* DNA is lower than 0.05 x SSC; 0.1 x SSC was shown to be a more acceptable level of stringency for the *Citrus* RFLP studies. The lower stringency tolerance of *Citrus* DNA may imply a relatively lower

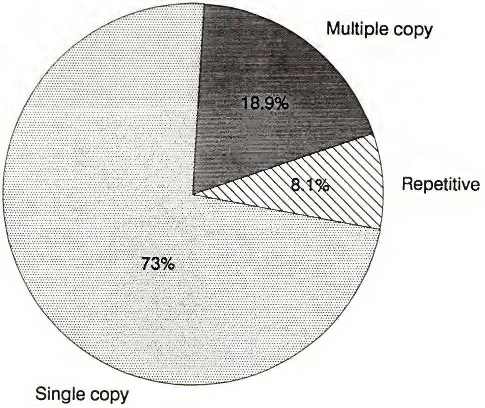
GC content in its genome than in rice or tomato, for example, because hybridization stringency is also dependent on the GC content of the genome, characteristic for each species (Goodenough, 1984).

Genome characterization by gene copy number determined with high stringency. Ninety seven clones selected by Procedure 5 were tested for clone copy numbers by Southern analysis with high stringency washes. Four clones showing maternal inheritance were excluded, so 93 clones were used to calculate percentages of single copy, multiple copy, and repetitive sequences. With high stringency washes, the *Pst*I genomic clones comprised 73% single copy, 18.9% multiple copy, and 8.1% repetitive sequences (Table 4-1 and Fig. 4-4). The copy number profile of the *Citrus Pst*I genomic library may be compared with the *Pst*I genomic library of rice and the random sheared library of tomato, two plants with genomes nearly the same size as *Citrus* (Guerra, 1984; McCouch et al., 1988; Zamir and Tanksley, 1988). The *Citrus* clones had a slightly lower percentage of single copy clones (85% and 78% in rice and tomato, respectively), a greater or equal percentage of multiple copy clones (12% and 18% in rice and tomato, respectively), and a greater percentage of repetitive clones (3% and 4% in rice and tomato, respectively) (McCouch et al., 1988; Zamir and Tanksley, 1988). The genomic clones used in this study were derived from *Pst*I digestion and likely represented the non-methylated portion of the *Citrus* genome. Moreover, non-methylated regions of plant genomes are thought to be enriched for single copy sequences, because they possibly represent coding regions (Burr et al., 1988; McCouch et al., 1988). One might expect, then, that the copy number profile of the *Citrus Pst*I genomic library might be skewed to underestimate the frequency of repetitive sequences and to overestimate single and multiple copy sequences in the *Citrus* genome. Therefore it may be risky to infer the gene copy number profile of the genome based on the

Table 4-1: Numbers (and percentages) of single copy, multiple copy, and repetitive clones obtained with high stringency washing.

Clone type	No. clones tested	No. (%) single copy	No. (%) multiple copy	No. (%) repetitive sequences
cDNA	19	12 (63.2)	6 (31.6)	1 (5.3)
genomic	74	54 (73.0)	14 (18.9)	6 (8.1)
Total	93	66 (71.0)	20 (21.5)	7 (7.0)

Fig. 4-4: Pie chart showing classification of *Pst*I genomic clones at high stringency (0.1 x SSC, 65°c). Results based on 74 clones.


Multiple copy
18.9%
8.1%
Repetitive
73%
Single copy

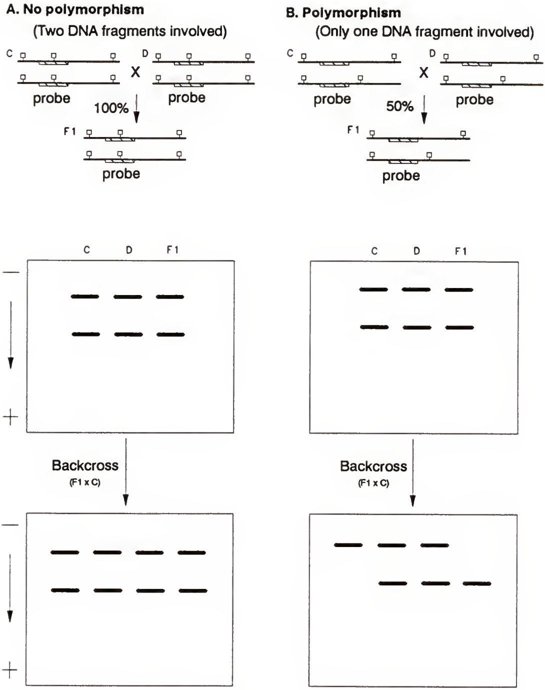
information derived from a *Pst*I-derived library. On the other hand, *Pst*I is invaluable in creating libraries rich in single copy clones.

It is interesting that the cDNA library studied was relatively rich in multiple copy sequences and had a lower percentage of single copy sequences compared with the *Pst*I genomic library (Table 4-1). The cDNA clones, being previously screened by their constructors, were not random clones and the number included in this study was not large enough to have confidence that observed percentages are true representative of the cDNA library. Nonetheless, the high percentage of multiple clones suggests that some of the structural genes of *Citrus* have more than one copy.

Restriction Fragment Length Polymorphism

Principles for identifying RFLPs in this study. Although RFLP markers provide significant advantages for gene mapping studies with tree species because of the large number of RFLPs that may be found in the same cross, the recognition of RFLPs from heterozygous parent-derived populations, like the population used in this research, requires care. Only clones showing putative Mendelian inheritance in 'Clementine' mandarin, 'Duncan' grapefruit, and their hybrid LB1-21 were thought to have potential for use in the linkage survey; this was the basic method used for the recognition of useful RFLPs. In Southern analysis, observing two bands in a lane does not insure heterozygosity for a locus, nor RFLP. Fig. 4-5 schematically represents a confusing situation that was encountered in this study. Thus, without further genetic analysis, it would be risky to conclude whether this banding pattern indicates allelic polymorphism. Eleven clones behaved in this fashion, and six of them revealed smaller band sizes (in kb) compared with the probe DNA length. This strongly suggests that at least 55% (6/11) of 11 clones

Fig. 4-5: Schematic representation of gels showing two situations where the same banding pattern would be obtained from the three individual plants tested. A. No polymorphism is involved. The probe hybridizes to the same two fragments in each individual. B. The enzyme/probe combination produces polymorphism. In this example, 50% of the F1 would have the same banding pattern as the parents. D = 'Duncan', C = 'Clementine', F1 = their hybrid LB1-21. Backcross progeny were derived from LB1-21 x 'Clementine'. Small squares represent sites where the DNA would be cut by the restriction enzyme used.


A. No polymorphism
(Two DNA fragments involved)
C
D
X
probe
probe
100%
F1
probe
C
D
F1
Backcross
(F1 x C)
B. Polymorphism
(Only one DNA fragment involved)
C
D
X
probe
probe
50%
F1
probe
C
D
F1
Backcross
(F1 x C)

are as the situation described in Fig. 4-5-A. For this reason, those banding patterns as described in Fig. 4-5 were not considered as polymorphisms.

Table 4-2 summarizes the results of the RFLP survey conducted in this study, including the polymorphism identified in different parent materials, with different enzymes and different library clones. Using 97 clones and six restriction enzymes for the survey, 68 clones (70.1%) revealed polymorphism among 'Clementine', 'Duncan', and their hybrid LB1-21; 59 clones (60.8%) were polymorphic between 'Clementine' and 'Duncan'; and 55 clones (56.7%) revealed polymorphism between 'Clementine' and LB1-21. The frequencies at which RFLPs were detected by different restriction enzymes were varible, with a tendency toward higher frequencies with *Eco*RI, *Bgl*II, *Eco*RV and *Hin*dIII, and lower frequencies with *Bam*HI and *Pst*I.

In most eukaryotes, only a very small part of the genome codes for proteins; the majority of DNA does not appear to serve a coding function (Klug and Cummings, 1986). Thus the *Pst*I-derived genomic clones may also represent non-coding areas of the genome, although they were thought to represent the coding area. Because the cDNA clones used in this study cannot be regarded as random clones (because they have previously been selected for being polymorphic in other situations by other researchers), the percentages of RFLP detected by them (Table 4-2) probably have been skewed to a higher level. However, the percentage of clones detecting polymorphisms was still greater in the genomic than the cDNA library. This was not surprising, because non-coding areas may be under less selection pressure in evolution than coding regions, thus allowing more variation in DNA sequences and giving a higher percentage of RFLPs.

Table 4-2: Polymorphisms detected by different restriction enzymes.

Restriction enzyme	genomic library: No. clones tested	genomic library: No. poly-morphic clones	genomic library: % (average)	cDNA library: No. clones tested	cDNA library: No. poly-morphic clones	cDNA library: % (average)
Among Clementine, Duncan and LB1-21 hybrid						
EcoRI	77	35	45.5	20	8	40.0
BglII	77	33	42.8	20	10	50.0
EcoRV	77	32	41.5	20	6	30.0
HindIII	34	13	38.2	7	1	14.3
BamHI	34	10	29.4	7	2	28.6
PstI	34	7	20.6	7	2	28.6
			(36.3)			(31.9)
any enzyme	77	55	71.4	20	13	65.0
Between Clementine and LB1-21 hybrid						
EcoRI	77	28	36.4	20	2	10.0
BglII	77	23	29.9	20	8	40.0
EcoRV	77	28	36.4	20	4	20.0
HindIII	34	14	41.2	7	0	0.0
BamHI	34	8	23.5	7	1	14.2
PstI	34	4	11.8	7	0	0.0
			(29.9)			(14.0)
any enzyme	77	46	59.7	20	9	45.0
Between Clementine and Duncan						
EcoRI	77	32	41.5	20	8	40.0
BglII	77	31	40.3	20	9	45.0
EcoRV	77	26	33.8	20	6	30.0
HindIII	34	12	35.3	7	1	14.3
BamHI	34	10	29.4	7	2	28.6
PstI	34	8	23.5	7	2	28.6
			(34.0)			(31.1)
any enzyme	77	49	63.6	20	10	50.0

Features of RFLP in Citrus: Evidence for Insertions and/or Deletions

Southern analysis showed that 58.2% (32/55) of the clones polymorphic between 'Clementine' and LBl-21 were polymorphic with more than one restriction enzyme. Burr and Burr (1983) reported that polymorphisms caused by insertions or deletions can be identified by the multiple effect they have when DNA is cleaved with two or more enzymes. RFLPs caused by insertion/deletion events have been found within domesticated "Corn Belt" maize germplasm (Helentjaris, 1987). McCouch et al. (1988) suggested that the lack of independence of probabilities for different enzymes detecting polymorphism in rice with a given probe was evidence for insertion/deletions, based on the principle that a base substitution would affect only the restriction pattern of one enzyme if the restriction sites for different enzymes along a piece of genomic DNA are independent and do not overlap.

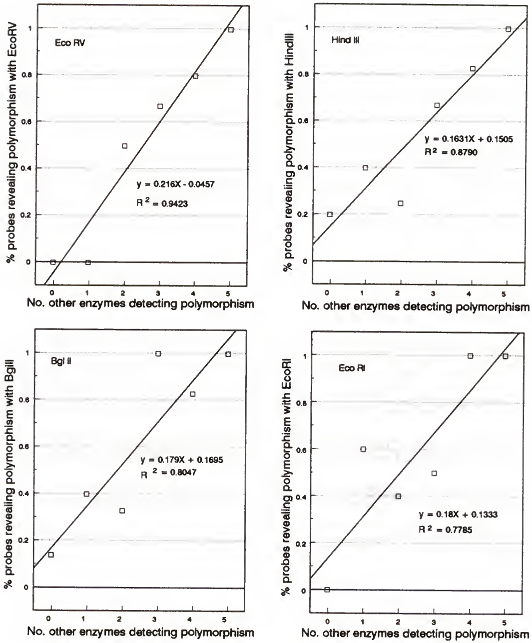
To study this phenomenon in *Citrus* in more detail, the relationship between the probability that a given enzyme detected polymorphism with a given probe in this study and the number of other enzymes detecting polymorphism with that same probe was plotted for regression. The results of Southern analysis from thirty nine clones tested with six different restriction enzymes were used for this study. Figure 4-6 shows the results of the linear regressions. The R^2 values ranged from 0.9423 to 0.5878, with three larger than 0.8 (*Eco*RV, *Hin*dIII and *Bgl*II), two greater than 0.7 (*Eco*RI and *Pst*I), and only *Bam*HI less than 0.6. The high R^2 values found with *Citrus* DNA clones implies that the probability of different enzymes detecting polymorphism with a given probe are not independent. If an RFLP results from an insertion or deletion, then RFLPs should be generated by any enzyme whose restriction site is near the affected area (Burr and Burr, 1983; McCouch et al., 1988). Based on this hypothesis, the results of regression (Fig. 4-6) would strongly suggest that a major mechanism for generating RFLPs in the *Citrus* genome is

insertion/deletion. The differences among R^2 values may be related to different restriction site frequencies of the different enzymes; high frequency restriction sites increase the frequency of restriction pattern changes involved in the insertion/delection events, and vice versa. Based on R^2 values, the order (from high to low) of restriction site frequency within the *Citrus* genome was thought to be: *Eco*RV, *Hin*dIII, *Bgl*II, *Eco*RI, *Pst*I and *Bam*HI.

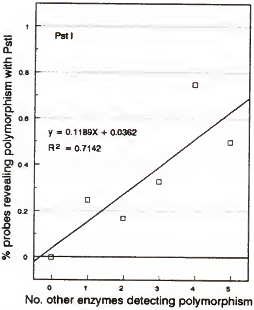
However, more extensive consideration of these results was warranted in light of the model described by McCouch et al. (1988). According to Burr and Burr (1983) and McCouch et al (1988), only those clones showing exactly the same banding patterns, as well as the same differences of band sizes with every polymorphism-detecting enzyme, can be satisfactorily explained as being caused by insertion/deletion events. Only one (pgCit015) of 32 clones in this study was of this type. Polymorphism of clone pgCit015 was caused by an insertion or deletion of approximately 1 kb and was detected by five of the six enzymes used in this study (photo not shown). The other clones that revealed polymorphism with more than one enzyme had different banding patterns with different enzymes (four of them are shown in Fig. 4-7), and cannot be satisfactorily explained by this model.

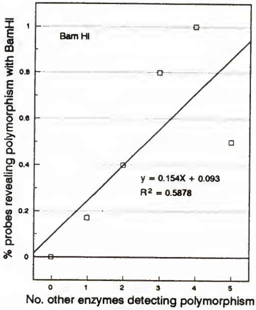
A more complex model or hypothesis to more satisfactorily explain the results of this study is presented in Fig. 4-8. This model is based on the hypothesis of McCouch et al. (1988) and is an attempt to explain the changes in banding patterns observed in this study with many of the clones. According to the model, the presence or absence of a restriction site for a specific enzyme on inserted or deleted fragments can cause changes of banding patterns, depending on the restriction enzyme under consideration. Figure 4-8 shows that both insertions and deletions may cause either increases or decreases of the restriction fragment length if there is a critical restriction site on the inserted or deleted fragment. Using this model,

Fig. 4-6: Plots derived from regressing the percentage of probes revealing polymorphisms detected by a specific enzyme vs. the numbers of other enzymes revealing polymorphisms with same probe.



Eco RV
y = 0.216X - 0.0457
R 2 = 0.9423
% probes revealing polymorphism with EcoRV
No. other enzymes detecting polymorphism
Hind III
y = 0.1631X + 0.1505
R2 = 0.8790
% probes revealing polymorphism with HindIII
No. other enzymes detecting polymorphism
Bgl II
y = 0.179X + 0.1695
R 2 = 0.8047
% probes revealing polymorphism with BglII
No. other enzymes detecting polymorphism
Eco RI
y = 0.18X + 0.1333
R 2 = 0.7785
% probes revealing polymorphism with EcoRI
No. other enzymes detecting polymorphism





(Fig. 4-6, continued)

Fig. 4-7: Southern analysis of some clones revealed polymorphisms with more than one enzyme, and different banding patterns with different enzymes. All of the blots were washed with high stringency condition (0.1 x SSC, 65°C). a) pgCit009; b) pgCit054L; c) pgCit026; d) pgCit071.

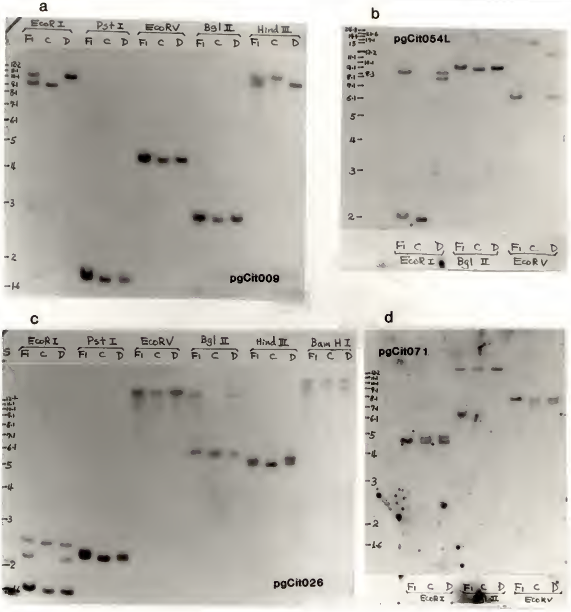
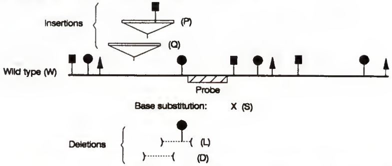
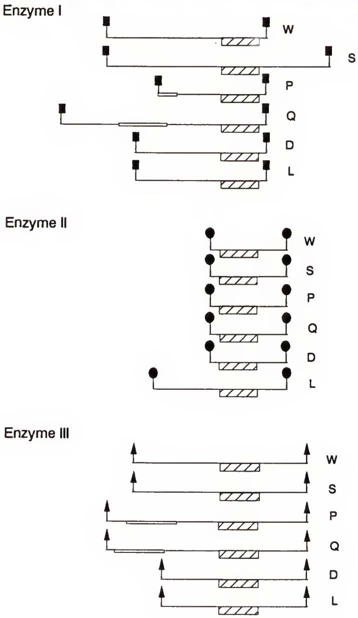

a
EcoR I
Pst I
EcoRV
Bgl II
Hind III
Fi C D
pgCit009
b
pgCit054L
Fi C D
EcoR I
Bgl II
EcoR V
c
EcoR I
Pst I
EcoRV
Bgl II
Hind III
Bam H I
Fi C D
pgCit026
d
pgCit071
Fi C D
EcoR I
EcoRV

Fig. 4-8: Schematic depicting the expected restriction fragments generated by different enzymes and different mechanisms including base substitution, insertion, and deletion. Note, base substitution can affect the restriction pattern for only one enzyme, whereas insertion/deletion is likely to affect more than one enzyme. Also, the insertion/deletion covering a restriction site can cause the change of restriction fragment length when comparing to the original fragment(w). The small squares, circles, and triangles represent the restriction sites of enzymes I, II, and III, respectively. W and S represent the original fragment (i.e. wild type) and the fragment caused by base substitution, respectively. P fragment is caused by an insert with a restriction site of enzyme I on it. Q fragment caused by an insert containing no restriction site. D fragment caused by a deletion containing no restriction site. L fragment caused by a deletion that covers a restriction site of enzyme II.

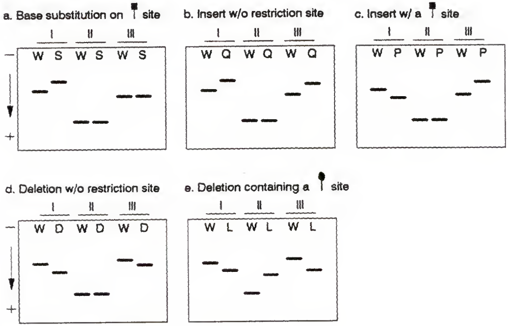
A. Mechanisms that cause size changes in restriction fragments.



B. Restriction fragment length changes caused by insertion, deletion, and basesubstitution.



C. Changes of banding patterns in different situations and with different enzymes.



(Figure 4-8, continued)

many unusual examples of polymorphism can be explained. For example, the RFLP detected by clone pgCit009 (Fig. 4-7-a) might be caused by an insert having an *Eco*RI or *Hin*dIII site on it, thus causing banding-style switching between these two enzyme restrictions. The plant F1 (LB1-21 hybrid) was heterozygous for pgCit009 with both *Eco*RI and *Hin*dIII, but homozygous in other three enzymes, while plants C ('Clementine') and D ('Duncan') were homozygous in all the five enzyme restrictions. To compare this real situation with the schematic model in Fig. 4-8-C-c, the enzymes I and III could be *Eco*RI and *Hin*dIII, and enzyme II could represent any one of *Pst*I, *Eco*RV, and *Bgl*II. The RFLPs detected by clone pgCit071 (Fig. 4-7-d) might be caused by an insert having an *Eco*RI or *Eco*RV site on it, thus causing switching in fragment sizes between these two enzyme restrictions. For this clone, the plant F1 is homozygous with all three enzyme restrictions, while both C and D are heterozygous in *Eco*RI and *Eco*RV restriction patterns but homozygous in *Bgl*II. Here enzymes I and III in Fig. 4-8-C-c could represent *Eco*RI and *Eco*RV in this real situation, and enzyme II could represent *Bgl*II.

At least eight different banding styles were observed among 17 polymorphic clones (Table 4-3), and that could be explained by this model. Seven of the eight styles contained more than one "exchangable" banding pattern within the style. More than half (53.1%) of the 32 clones where polymorphism was revealed by more than one enzyme could easily be explained by this model based on situations where two or three alleles are involved (Table 4-3). Of all polymorphism-detecting clones, 30.9% (17/55), can be easily explained by this insertion/deletion model (Tables 4-3 and 4-4). The other 15 of the 32 clones also could be explained with this model, although the explanations become more complex because of three or more alleles being involved within these three *Citrus* lines.

Table 4-3: The classification of banding styles found in this study and the percentage of clones in each class. The different banding patterns within each style were observed with different restriction enzymes and thought to be caused by the existence of a restriction site on the insertion/deletion fragment. Each set of banding patterns is in the order of LB1-21 hybrid, 'Clementine' mandarin, and 'Duncan' grapefruit, as in the gels shown in Figs. 4-2, 4-3, and 4-7.

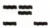
Style [1]	No. clones	% (total: 55)	Clone
A. Two alleles involved.			
(a) 	6	10.9	pgCit005, 026, 016, 037, 069 and pcPt001
(b) 	4	7.3	pgCit015, 020, 056, and 071
(c) 	2	3.6	pgCit009, 021
(d) 	1	1.8	pgCit027
(e) 	1	1.8	pRLC91
Subtotal	(14)	(25.5)	
B. more than two alleles involved.			
(f) 	1	1.8	pgCit019
(g) 	1	1.8	pgCit054L
(h) 	1	1.8	pgCit018
Subtotal	(3)	(9.4)	
Total	17	30.9	

Table 4-4: The fitness of the insertion/deletion model for the clones revealing RFLP by more than one enzyme.

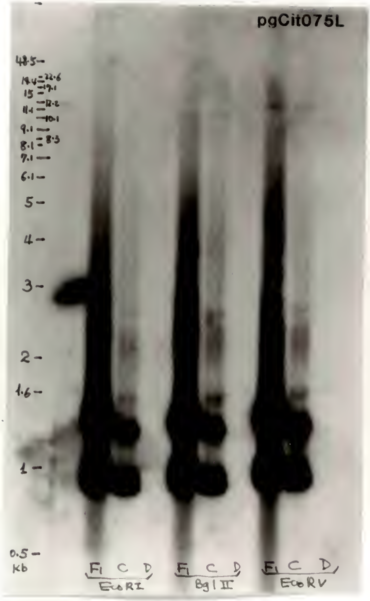
Description	No. clones	percentage (divided by A)	Percentage (divided by C)
A. Total no. polymorphic clones	55	100.0	-
B. Clones revealing polymorphism with only one enzyme	23	41.8	-
C. Clones revealing polymorphism with more than one enzyme	32	58.2	100.0
D. Polymorphism explainable by insertion/ deletion model	17	30.9	53.1

It has been found in humans and *Drosophila* that multiple single-base mutations occur at five to ten percent of polymorphic restriction sites assuming that enzymes recognizing 6-base sites were used on samples of 50 to 100 individuals (Hudson, 1982, 1989). In this study, at least 18 polymorphic clones appeared to recognize loci where more than two alleles were present in the banding patterns of the three plants surveyed (The banding patterns of three of the clones are diagrammed in Table 4-3-B). The polymorphisms detected by these 18 clones should have resulted from multiple mutations, supposing that all *Citrus* species have a common ancestor. Thus the frequency of multiple mutations detected in this study can be estimated as 32.7% (18/55), which is even higher than humans and *Drosophila*. *Citrus* plants are highly heterozygous and genetically unstable (Soost and Cameron, 1975). A high frequency of insertion/deletions may have played an important role in this hetero-zygosity and genetic instability.

A Clone That May Imply the Existence of a Transposon by Its Unusual Banding Pattern in Southern Analysis

Figure 4-9 shows the result of Southern analysis using clone pgCit075L as a probe. There were three characteristics that made this clone worthy of special notice. First, it exhibited materal inheritance, because there was no signal in any of the lanes containing the DNA of the male parent ('Duncan' grapefruit), but the same signals in all of the lanes of the female parent and the hybrid tree ('Clementine' mandarin and LB1-21, respectively). Second, all the lanes of the female parent and hybrid tree showed a smeared signal, indicating a repetitive fragment. Third, there were two dark bands at ca. 0.9 and 1.4 kb positions in all of the female parent and hybrid lanes. The strong signal of these two bands was indicative of a large number of DNA copies at these points.

Fig. 4-9: The result of Southern analysis using clone pgCit075L as a probe. The high stringency condition (0.1 x SSC, 65°C) was used for washing.


pgCit075L
48.5
19.4
22.6
15
17.1
11.1
12.2
10.1
9.1
8.1
8.3
7.1
6.1
5
4
3
2
1.6
1
0.5
kb
F1 C D
EcoRI
F1 C D
BglII
F1 C D
EcoRV

Because all of the DNA used in Southern blotting was prepared by the modified Dellaporta procedure as described in Chapter 3, the preparations should include nuclear, mitochondria, and chloroplast DNA. The results indicated that there were no homologous DNA sequences between probe and the 'Duncan' male parent (in either the nuclear or organellar genomes), whereas repetitive sequences existed in both the female parent (Clementine) and the hybrid (LB1-21) that were homologous to probe DNA. Further, because smeared signals are not a feature of organellar genomes, it is likely that the homologous sequences reside inside the nucleus and not in the organelles. The two dark bands with uniform sizes in all of the lanes containing DNA of female and hybrid trees restricted by different enzymes suggested two possibilities: they might come either from a high number of repeated sequences inside the nucleus, or alternatively from a plasmid-like element inside the mitochondria. However, the first possibility can be excluded because of the maternal inheritance. Therefore, it is possible that these two dark bands might come from a circular episomal DNA and its concatemer. It was possible for such an episomal DNA to be cloned into the vector plasmid during the construction of the *Pst*I-derived library, if there is a *Pst*I restriction site on it. If this is true, then the molecular weights of the episomal DNA and the probe (i.e. the clone pgCit075L) will most likely be the same. This appears to be true in this study, because both of them were estimated to be ca. 1.2 kb (see Appendix 1 and Fig. 4-9. Circular molecules usually migrate faster in an agarose gel, so the band at the 0.9 kb position was thought to be ca. 1.2 kb in molecular weight). The reasons for this hypothesis are given below.

Plasmid-like DNA molecules of low molecular weight have been found in mitochondria of many different species including *Zea mays* (Kemble and Bedbrook, 1980; Kemble et al., 1983), *Sorghum* (Chase and Pring, 1985, 1986), *Brassica*

(Palmer et al., 1983), *Vicia faba* (Goblet et al., 1983; Wahleithner and Wolstenholme, 1987); *Oryza sativa* (Kadowaki et al., 1988); and *Neurospora* (Nargang et al., 1984). They can be circular or linear (Kemble and Bedbrook, 1980, Nikiforova and Negruk, 1983, Chase and Pring, 1985, 1986). It is possible that similar DNA molecules can also be found in *Citrus*. If this hypothesis is true, the molecules may have resided in many (but not all) *Citrus* cultivars, because the DNA clone was derived from a cultivar ('Temple' tangor) other than any of the materials used in Southern blotting. Moreover, Kemble et al. (1983) reported that, in maize, sequences homologous to episomal mitochondrial DNAs could be found in the nuclear genome, indicating that the sequences are transported from the mitochondrion to the nucleus. Nargang et al. (1984) concluded that the DNA sequence and genetic organization of a *Neurospora* mitochondrial plasmid suggested a relationship to introns and mobile elements. These properties would make the smear signal of the female and hybrid lanes in Fig. 4-9 understandable. Therefore, it was suggested that the DNA sequence in clone pgCit075L may be related to sequences of a transposon.

A future experiment can be designed to prove/disprove this hypothesis. First, other plant filters may be necessary for confirming maternal inheritance, which is one of the key points for building this hypothesis. An ideal plant filter should comprise two *Citrus* lines (as parent trees) and their hybrid(s). The female parent tree should be able to give hybridizing band(s) or signal in Southern analysis when using this clone as a probe. Ideal candidates for use as a female parent include 'Clementine' mandarin (proven in this study) and 'Temple' tangor (the source of the library clones). The ideal male parent tree should show somewhat different banding patterns compared with the female parent. Second, the *Pst*I-digested DNAs, including all of the *Citrus* lines in the plant filter, need to be included in the Southern blot. Such a blot could be used to examine the circular

episome hypothesis. If the clone came from a circular episome, the banding pattern revealed by the *Pst*I-digested DNA should be different from DNA digested with other enzymes. In the lanes of *Pst*I-digested DNA, only one unique band is expected at the 1.2 kb position, if the *Pst*I restriction is complete. For further confirmation, it would be advisable to have a lane containing only the pgCit075L clone DNA on the blot for comparing with the band produced by the *Pst*I-digested DNA. Third, a subcloning of the clone pgCit075L would be necessary, because this DNA clone comes from an *E. coli* colony (named as FA9 in the microtiter tray) that harbors two *Citrus* DNA clones. The subcloned pgCit075L will be useful for DNA sequencing and further studies.

Some Conclusions About the RFLP Study in Citrus

Based on the experiments performed in Chapter 3, a much more detailed survey of the library was performed in this chapter. General characteristics of the *Citrus* genome, including gene copy profile, the extent of DNA methylation, and the major mechanism causing RFLPs, were also studied. The high percentage of single copy clones in the *Pst*I-derived library (Fig. 4-4) implies that the *Pst*I enzyme is invaluable for library construction in *Citrus* RFLP studies. It was concluded that insertion/deletion is a major mechanism causing RFLP in *Citrus*. Also, the Southern analysis of a clone (pgCit075L) strongly implied a transposon-related sequence in *Citrus*, and some strategies for the further studies of this hypothesis were suggested.

CHAPTER 5

GENE MAPPING IN CITRUS USING RFLP AND ISOZYME ANALYSIS

Introduction

The difficulties encountered when conventional plant breeding methods are employed for *Citrus* variety improvement were discussed in Chapter 1, along with the need for an alternative approach to enhance breeding efficiency, i.e. a linkage map. This chapter describes experiments, based on the results obtained in Chapter 4, that were designed to analyze linkage relationships among selected genomic and cDNA clones and isozyme markers so that a linkage map could be constructed. Also, an experiment was designed to estimate the *Citrus* genome size through the mapping process using different numbers of markers. Transposon activity was proposed to interpret the unusual phenomenon of progeny analysis for some clones. The mapping results based on this assumption were then examined. A discussion of the general features of the linkage map and the estimate of the *Citrus* genome size conclude this chapter.

Materials and Methods

Plant Population Used for Progeny Analyses

A backcross progeny of 65 hybrid seedlings was used for RFLP and isozyme analyses to produce segregation data. The 65 plants were randomly selected from a population totaling 192 plants that was derived from the backcross of LB 1-21 ('Clementine' mandarin x 'Duncan' grapefruit) x 'Clementine' mandarin, made by Dr.

Fred G. Gmitter at CREC in 1986 (cross number #8610 in the breeding program). The backcross population displayed a wide range of variation for morphological characters such as leaf size and shape, tree growth habit, vigor, etc., indicating substantial genetic variation in this population. All plant materials were maintained at CREC, Lake Alfred in pots in the greenhouse and were under intensive care. They were in good health when leaves were collected for DNA isolation and isozyme analysis. Young leaves were harvested before full expansion for DNA isolation using the modified Dellaporta procedure described in Chapter 3. Fully expanded adult leaves were used for isozyme analysis.

Isozyme Survey

To look for segregating isozyme loci in the backcross progeny, the female parent (LB 1-21), male parent ('Clementine' mandarin), and the donor parent ('Duncan' grapefruit) were assayed for allozymic differences of different enzymes, including: phosphoglucomutase (PGM, E.C.2.7.5.1), malate dehydrogonase (MDH, E.C.1.1.1.37), phosphoglucoisomerase (PGI, E.C.5.3.1.9), 6-phosphogluconate dehydrogenase (6-PGD, E.C.1.1.1. 49), shikimate dehydrogenase (SDH, E.C.1.1.1.25), glutamate-oxaloacetate transaminase (GOT, E.C.2.6.1.1), peroxidase (PER, E.C.1.11.17), malic enzyme (ME, E.C.1.1.4.0), leucine amino peptidase (LAP, E.C.3.4.11.1), acid phosphatase (ACP, E.C.3.1.3.2), isocitrate dehydrogenase (IDH, E.C.1.1.1.42), superoxide dismutase (SOD, E.C.1.15.1.1), and alkaline phosphatase (AKP, E.C.3.1.3.1), that cumulatively represent at least 20 loci (Torres, 1983; Roose, 1988; F. G. Gmitter, unpublished data).

Starch gel electrophoresis and enzyme staining were according to Gmitter (1985) and Durham et al. (1987) with modifications. Constant current of 60 mA/gel was used instead of constant voltage. With constant current, the gels frequently

became too hot. To solve this problem, the gel was run in the cold room (5°C), and two frozen gel packs were placed on top of the starch gel to remove heat.

Of the 13 isozymes surveyed, 6 isozymes representing 7 loci were expected to be segregating in the progeny. These loci were *sdh, pgm-2, got-1, got-2, pgi, acp* and *idh*. Segregation data for each isozyme were collected from the 65 progeny plants, except got-2 was unreadable in most of the gels for progeny analysis. The isozyme analyses revealed no evidence of outcrossing in the progeny, thus indicating the suitability of this population for linkage study.

Collecting RFLP Segregation Data

From the results of Southern analysis performed in Chapter 4, clones capable of showing simple, clear-cut band(s) and appearing to follow Mendelian inheritance, as indicated by examination of patterns from the two parent trees and 'Duncan' grapefruit, were selected as candidate clones for progeny analysis. Clone/ endonuclease combinations revealing polymorphism between the female parent (LB 1-21) and male parent ('Clementine' mandarin) of the back-cross progeny were used in progeny analysis. There were 55 clones showing this capability, including 46 genomic clones and 9 cDNA clones.

Some clone/endonuclease combinations produced exactly the same banding pattern consisting of two clear-cut bands in both female and male parents of the backcross progeny. A 1:2:1 segregation ratio in the progeny could be expected if the two-band pattern resulted from heterozygosity at the locus. These results (1:2:1) were not always observed, because a homozygous plant is also capable of producing double bands when the homologous probe sequences overlap with two adjacent fragments, as discussed in Chapter 4. Only 5 of these clones, for which

fragment size measured by the DNA size markers in the gels was larger than that of probes, were used in progeny analysis.

The methods for restriction digests, electrophoresis, and Southern analyses performed with progeny plants were the same as described in Chapter 3, except that only the important band position areas in the agarose gels were excised and used for Southern blotting. In most cases, all of the progeny trees (65 plants usually from 3 gels) were arranged on the same blot. The progeny blots used in Southern analysis also included the female parent (LB1-21), male parent ('Clementine' mandarin), and the donor parent ('Duncan' grapefruit) to double check the inheritance patterns and to determine the banding patterns of progeny trees. The washing stringency for every clone was the same as that used for detecting the polymorphism in that clone.

Linkage Analysis and Map Construction

The entire set of markers (including RFLP and isozyme markers) was processed through the LINKAGE-1 computer program for linkage analysis (Suiter et al., 1983). Chi-square goodness-of-fit values for each segregating locus, independent assortment deviation of all pairs of segregating loci, and the maxmum likelihood estimates of recombination-frequency between linked loci were calculated on an IBM personal computer with the LINKAGE-1 program. Three different significance levels (0.05, 0.01 and 0.001) were used to determine linkages. Standard errors for each calculated recombination-frequency were also given by the program. Missing data were tolerated to some extent by the program. This program was capable only of performing two-point analysis, so the linkage map was deduced as the best fit to these values and drawn by hand following the principle of the three-point-cross method (Suzuki et al., 1980). To do this, linkage groups comprising

three loci each were selected first. For each possible gene order in the group (such as "a-b-c"), the distances between each gene pair (e.g. ac, bc and ab) were calculated and compared with the computer printout for the best fit to the computer data.

Estimation of Genome Size

The total number of centimorgans in the genome of an organism can be estimated from partial linkage data as follows. Let G denote the length of the genome in cM (i.e. genome size); let the function $P(N)$ denote the percentage of the determined linked markers, where N is the total number of random markers used for linkage analysis; let the function $D(N)$ denote the average distance of the adjacent loci in the linkage map constructed by using N markers for linkage analysis. Thus the coverage (in cM) of the map derived from these N markers can be calculated as $N \times P(N) \times D(N)$. If m is the random sample size that makes $P(m) = 100\%$, it means that the genome is saturated with mapped loci, and for every location in the genome there will be at least one mapped locus close enough to be detected by linkage analysis. Then, an estimate of G is given by the equation $G = m \times P(m) \times D(m) = m\ D(m)$. It only remains to determine the functions $P(N)$ and $D(N)$.

These two functions can be determined by mathematical regressions. Each regression uses a series of data (i.e. the percentages of mapped markers in the samples and the average distances of loci in the maps) derived from a series of random samples (marker groups) comprising different numbers of markers. The series of sample sizes used for linkage analysis must cover a vast extent of interval along the X-axis. All of the available markers in every random sample used for linkage analysis are informative of the same number of meioses (i.e. all of the segregation data are from the same population). The sample sizes are plotted to the

percentages of mapped makers and the average distances of loci for regressions. The precision of the estimate of genome size depends first upon the reliability of the regressions, and second, the degree of extrapolation necessary for the estimation.

Results and Discussion

Segregating RFLP and Isozyme Loci

Forty two segregating loci, including 36 RFLP loci and 6 isozyme loci, were included in this study (Tables 5-1 and 5-2). All of the RFLP variants were detected with high stringency, except two loci (pgCit020 and 041) that were detected with moderate stringency conditions. According to the parental genotype, 34 loci were expected to segregate in a 1:1 ratio, 6 loci were expected to segregate in a 1:2:1 ratio, and 2 loci were expected to segregate in a 1:1:1:1 ratio. Among these 42 loci, only 9 loci were homozygous in female parent (LB1-21 hybrid), while 25 loci were homozygous in male parent ('Clementine'). Since a homozygous genotype cannot be used to determine recombination, most contributions of detectable recombination in this linkage study were from the female gamete. There were 3 RFLP loci, i.e. pgCit009, pgCit046 and pgCit048, that showed extra alleles in the progeny analysis that were not expected from the parental genotypes (Fig. 5-1); these will be discussed in the following sections.

There were 11 RFLP loci (26.5% of the total, and 30.6% of the RFLP loci) that deviated significantly from the expected monogenic ratios, at the 0.05 level. None of the isozyme loci deviated from expected ratios at the 0.05 level of significance. There was one locus (*idh*) deviated at 0.1 level (P value = 0.054). It is of interest to note that in an intergeneric backcross population derived from *Citrus grandis* and *Poncirus trifoliata*, there were higher percentages of distorted

segregation (37% and 56%) in RFLP and isozyme locirespectively (R. Durham, personal communication). Although the reason for segregation distortion is unknown, it has been noted in other inter- and intraspecific populations including *Lycopersicon* (Bernatzky and Tanksley, 1986; Helentjaris et al., 1986; Paterson et al., 1988),*Capsicum annuum* (Tanksley, 1984), *Oryza sativa* (McCouch et al., 1988), *Solanum tuberosum* (Bonierbale et al., 1988), and *Zea mays* (Helentjaris et al., 1988). The percentages of loci showing distorted segregation ranged from 12% (as in an *Oryza sativa* intraspecific population, McCouch et al., 1988) to 68% (as in a *Lycopersicon* interspecific population, Paterson et al., 1988). In general, the percentage of loci showing distorted segregation observed in this study is consistent with that observed in other genera and species.

Among the 11 loci showing distorted segregation, 7 were expected to segregate in a 1:1 ratio, and their deviation could be examined. Table 5-3 showed that 4 of these 7 loci had an excess of alleles contributed by the female parent and a deficiency of male alleles in the progeny; the other 3 loci deviated in the other direction. Thus there was no evidence of preferential deviation toward either parental genotype. This result was consistent with previous research in *Citrus* where random isozyme segregation distortion has been observed (Torres et al., 1985).

Unusual Banding Patterns in Some Clones Might Be Caused by Transposon Activity

Figure 5-1 shows results of progeny analysis with different clones. Three clones, pgCit009, pgCit046, and pgCit048, produced banding patterns that did not follow Mendelian segregation. In these cases, some progeny trees in the population showed extra bands that were not expected based on the parental genotypes; 43%

Table 5-1: Segregating RFLP loci detected in progeny analyses.

Clone	Enzyme	Genotype[1] F	Genotype[1] M	Expected ratio	Observed ratio	Chi-square
pgCit001L	EcoRV	ab	aa	1:1	aa(32): ab(33)	0.02
005	EcoRI	ab	aa	1:1	aa(28): ab(36)	1.00
009	EcoRI	ab	aa	1:1**	-	-
010	EcoRV	aa	ab	1:1	aa(34): ab(30)	0.25
011	EcoRV	aa	ab	1:1	aa(24): ab(41)	4.45*
015	BglII	aa	ab	1:1	aa(42): ab(23)	5.55*
016	EcoRI	ab	aa	1:1	aa(28): ab(37)	1.25
017	EcoRV	ab	aa	1:1	aa(27): ab(31)	0.28
019	EcoRI	ab	ac	1:1:1:1	aa(25): ab(9): ac(18): bc(13)	8.78*
020	BglII	ab	aa	1:1	aa(40): ab(20)	6.67*
021	EcoRI	ab	aa	1:1	aa(33): ab(31)	0.06
025	EcoRV	ab	aa	1:1	aa(41): ab(24)	4.45*
026	EcoRI	ab	aa	1:1	aa(35): ab(30)	0.38
027	EcoRI	aa	ab	1:1	aa(31): ab(34)	0.14
028	BglII	ab	aa	1:1	aa(32): ab(27)	0.42
031S	BglII	ab	aa	1:1	aa(31): ab(32)	0.02
035	BglII	ab	aa	1:1	aa(32): ab(33)	0.02
037	EcoRI	ab	aa	1:1	aa(26): ab(32)	0.62
041	EcoRV	ab	aa	1:1	aa(31): ab(34)	0.14
046	EcoRV	ab	aa	1:1**	-	-
048	EcoRI	ab	aa	1:1**	-	-
049	EcoRI	ab	aa	1:1	aa(32): ab(33)	0.02

(Table 5-1, continued)

Clone	Enzyme	Genotype[1] F	M	Expected ratio	Observed ratio	Chi-square
pgCit053	BglII	ab	ab	1:2:1	aa(10): ab(36): bb(19)	3.25
054L	EcoRI	ab	aa	1:1	aa(33): ab(32)	0.02
054S	EcoRV	ab	aa	1:1	aa(31): ab(34)	0.14
056	EcoRI	aa	ab	1:1	aa(41): ab(24)	4.45*
062	EcoRI	aa	ab	1:1	aa(48): ab(17)	14.78*
069	BglII	ab	aa	1:1	aa(26): ab(36)	1.61
071	EcoRV	aa	ab	1:1	aa(51): ab(14)	21.06
074	EcoRI	ab	aa	1:1	aa(35): ab(29)	0.56
pcPt001	EcoRI	ab	aa	1:1	aa(27): ab(38)	1.86
pRLc24	BglII	aa	ab	1:1	aa(42): ab(23)	5.55*
27L	BglII	ab	ab	1:2:1	aa(24): ab(31): bb(10)	6.17*
32	EcoRV	ab	ab	1:2:1	aa(9): ab(56): bb(0)	36.48*
89	EcoRV	ab	ab	1:2:1	aa(11): ab(43): bb(11)	21.38*
91	EcoRV	ab	ab	1:2:1	aa(19): ab(31): bb(15)	0.63

1 F and M denote female parent (LB1-21 hybrid) and male parent ('Clementine' mandarin), respectively. He and Ho denote heterozygous and homozygous, respectively.
* Segregation distortion indicated at 0.05 level in chi-square test.
** With non-parental alleles in some progeny trees.

Table 5-2: Segregating isozyme loci detected in progeny analyses.

Isozyme	Parental Genotype Female	Parental Genotype Male	Expected ratio	Observed ratio	Chi-square	P value
ACP	FM	MM	1:1	FM(33): MM(28)	0.41	0.522
GOT1	FS	SS	1:1	FS(26): SS(39)	2.60	0.107
IDH	FF	FM	1:1	FF(38): FM(23)	3.69	0.054
PGM1	IS	IF	1:1:1:1	FI(22): FS(11): II(13): IS(19)	4.85	0.183
PGI	FS	FS	1:2:1	FF(16): FS(36): SS(12)	1.50	0.472
SDH	IS	II	1:1	II(31): IS(34)	0.138	0.709

Table 5-3: Loci that were expected to segregate in an 1:1 ratio but showed segregation distorted toward one of the parental genotypes.

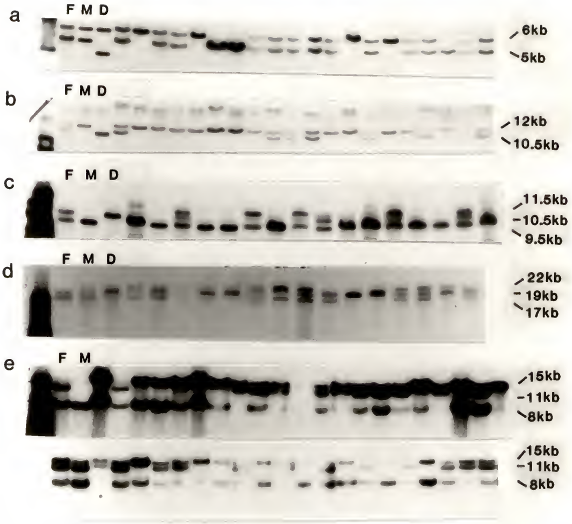
Clone/enzyme	Off. genotype[1] F	M	P value	Skew[2]
pgCit001/EcoRV	24	41	0.034980	m
015/BglII	42	23	0.018440	f
020/EcoRV[3]	20	40	0.009823	m
025/EcoRV	24	41	0.034980	m
056/EcoRI	41	24	0.034980	f
062/EcoRI	48	17	0.000121	f
pRLc24/BglII	42	23	0.018440	f

1 There are only two different genotypes in the offspring: F is the same as female parent (LB1-21 hybrid); M is the same as male parent ('Clementine').

2 m, skewed toward male parental genotype (there were more male genotype than femal genotype in the progeny) ; f, skewed toward female parental genotype (more female genotype than male genotype in the progeny).

3 Polymorphism was detected by moderate stringency.

Fig. 5-1: Southern analysis of parent trees and part of the progeny population, with clones pgCit035 (a) and pgCit053 (b) showing normal banding patterns as expected from the parental genotypes; whereas clones pgCit009 (c), pgCit046 (d), and pgCit048 (e) showing unusual banding patterns in some of the progeny plants that were not expected from parental genotypes. In the pictures, F, M, and D denote the female tree (LB1-21 hybrid), male tree ('Clementine'), and donor tree ('Duncan'), respectively. The 11.5, 22, and 15 kb alleles revealed by clones pgCit009, pgCit046, and pgCit048, respectively were suggested to be analogous to the alleles **a** depicted in the Fig. 5-2. And the 10.5, 19, and 8 kb alleles revealed by the same 3 clones were suggested to be analogous to the alleles **a*** in Fig. 5-2.



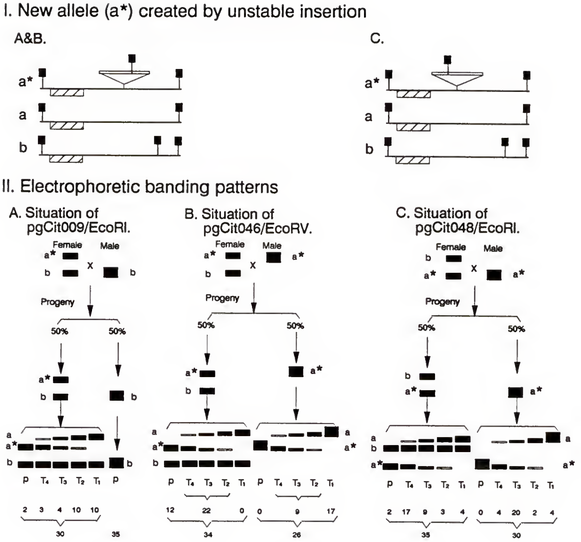
a
F M D
6kb
5kb
b
F M D
12kb
10.5kb
c
F M D
11.5kb
10.5kb
9.5kb
d
F M D
22kb
19kb
17kb
e
F M
15kb
11kb
8kb
15kb
11kb
8kb

(28/65), 80% (48/60) and 97% (63/65) of the progeny had extra bands revealed by pgCit009/*Eco*RI, PgCit046/*Eco*RV, and pgCit048/*Eco*RI, respectively (Fig. 5-2).

Several possibile explanations were considered for these unexpected results. First, the unusual bands might have resulted from incomplete digestion. This seemed not to have been the case, however, because the extra bands in different plants were quite uniform in both shape and size when the same clone was used. Second, the unusual restriction patterns might have been the result of methylation, especially if plants were under stress or in different developmental stages. However, there is little support for this mechanism because all plants were of the same age and were growing in healthy condition. Third, the extra bands might have been caused by two overlapping loci; this possibility also can be excluded, because there were complementary dosage effects between the extra bands and some expected bands (i.e. the bands analogous to the schematic band a* in Fig. 5-2). Based on the observation of the band dosages, it was concluded that exchangeability was the rule between some of the bands. Therefore a hypothesis of transposon activity, as schematically depicted in Fig. 5-2, was suggested to explain these phenomena.

As discussed in Chapter 4, many of the *Citrus* RFLPs might have been caused by insertions or deletions. Among these 3 unusual clones, pgCit009 was assumed to be detecting an insertion or deletion, with an *Eco*RI or *Hin*dIII site on the insertion/deletion fragment as discussed in Chapter 4 (see Table 4-3, Fig. 4-6, and Fig. 4-7). The other two clones showed RFLP only with one out of three enzymes used in the RFLP survey; therefore the information from these 2 clones was considered suggestive, but not conclusive, of the insertion/deletion model suggested in Chapter 4. Clearly, the possibility of insertion/delection cannot be ruled out for these two clones.

Fig. 5-2: Schematic models depicting the expected banding patterns based on the assumption of transposon activity, and showing possible explanations for the results of progeny analyses of clones pgCit009, pgCit046, and pgCit048. Note: All migration of DNA fragments in the gels is from top to bottom in the diagram. **a*** is the allele **a** containing an unstable insertion. **P** designates the parental genotypes. **T** designates the genotypes caused by reversional excision of the insertion in allele **a***. In **T1**, excision occurred at an extremely early stage in one cell, and that cell finally developed into the whole tissue from which DNA was extracted and leading to the total loss of band **a*** in the gel. Later times of excision, leading to tissue with a mixture of cell types, are designated by greater **T** numbers. The numbers below the picture are the actual number of trees showing the various genotypes in Southern analysis.



I. New allele (a*) created by unstable insertion
A&B.
C.
a*
a
b
II. Electrophoretic banding patterns
A. Situation of pgCit009/EcoRI.
B. Situation of pgCit046/EcoRV.
C. Situation of pgCit048/EcoRI.
Female
Male
X
Progeny
50%
50%
P T4 T3 T2 T1 P
2 3 4 10 10
30
35
12 22 0
34
0 9 17
26
2 17 9 3 4
35
0 4 20 2 4
30

If an insertion is reversible and excision occurred independently in different somatic cells, the DNA blot prepared from that somatic tissue would give unusual banding patterns. Further, the time at which the reversional excision occurs will determine the banding patterns. During the development of somatic tissue, an extremely early reversional excision in one cell that finally developed into the whole tissue would lead to the complete disappearance of one of the bands expected from the parental genotypes (i.e. band a^* in Fig. 5-2) and would create a new band (i.e. band a). No excision, or an extremely late reversional excision, would preserve the expected banding pattern and create no new band. Reversions that occurred between these two extreme situations will produce 3 bands (a, a^* and b bands in Fig. 5-2) with different dosage effects between 2 of them (a and a^*) depending on the time of excision. The time when a reversion occurred can be estimated by studying the band dosages and patterns. Based on this model, the extra bands found in some of the progeny plants were thought to be strongly suggestive of transposon activity.

When the "normal" genotypes (without excision, only parental alleles appear) for these three loci were inferred for each tree based on the proposed model (Fig. 5-2), chi-square tests showed a good fit to the expected segregation ratio for each case (Table 5-4). These results support the hypothesis of transposable insertion and excision proposed to interpret and explain the data. Further, two of these three clones (pgCit009 and 048) also have been used in another mapping study with a different progeny population derived from *Citrus* and citrumelo (hybrid of *C. grandis* and *Poncirus trifoliata*). In that study, the results of progeny analysis with these two markers revealed no extra bands and inheritance was as predicted by Mendelian segregation (R. Durham, unpublished data.) The different behavior of these markers in different progeny implies that the occurrence of extra alleles in different

Table 5-4: The chi-square test for the fitness of expected segregation ratios of 3 clones showing unusual banding patterns in progeny analysis. The genotypes were assigned based on the assumption of tranposon activity as depicted in Fig. 5-2.

Clone	Genotype aa	Genotype ab	Total no.	Exp. ratio	Chi-square	p value
pgCit009	30	35	65	1:1	0.3846	0.54
pgCit046	34	26	60	1:1	1.0600	0.30
pgCit048	35	30	65	1:1	0.3846	0.54

populations is genotype dependent, and probe independent. This implication is consistent with the transposon phenomenon and further supports the hypothesis.

Another question must be addressed here: why were these transposable insertions not excised in the tissues (of parent trees) that were collected for DNA isolation? The answer to this question is critical, because a transposon might not be moving all the time. Moreover, a transposon can be autonomous or non-autonomous for insertion/excision activity, such as the well studied Ac-Ds system in maize (Fedoroff, 1983; 1984). It is possible that some transposons in *Citrus* are non-autonomous. In that case, the transposable activity will be genotype dependent, and an activating factor will be needed for its activity. Thus, in genomes without activating factors, the alleles caused from insertions will become stable. Hake et al. (1989) reported that the Ds insertion in maize could create banding patterns similar, in their progeny population, to those shown by clones pgCit009, pgCit046, and pgCit048 (Fig. 5-1), with Ac introduced into the genome as the activator. In that study, the progeny genome without the activator (i.e. Ac element) showed only parental alleles in Southern analysis, and only some of the plants with activator introduced into their genomes showed extra alleles. Their results suggested that only the existence of an activator inside the genome can create the probability of reversional excision for the non-autonomous element; however, excision may not necessarily occur, even if the activator is present. The same principle may also prove to be involved with the materials used in this study, where the activating factor may come from the genome (non-autonomous) or from the insert itself (autonomous).

In this study, the frequencies of plants with extra bands detected by clones pgCit009/*Eco*RI, 046/*Eco*RV, and 048/*Eco*RI were quite different, thus implying that transposable activities were different in these three cases, according to the

transposon model suggested in Fig. 5-2. Besides, clones pgCit009 and pgCit048 detected transposable activity with the same enzyme (*Eco*RI), but inferred different insert sizes from the assumed inserted band and the extra band (i.e. the differences of sizes between the a- and a^*-homologous band in Fig. 5-1). If transposon activity occurs in *Citrus*, there must have been more than one transposon involved, based on the size differences of the identical insert.

Mapping Clones Having Unusual Banding Patterns Based on Assumption of Transposon Activity

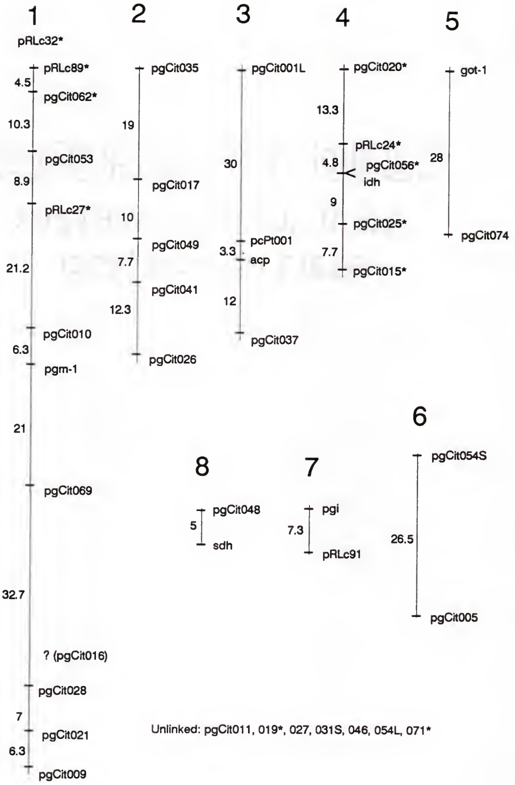
Based on the proposed transposon hypothesis, the "normal" genotypes for every progeny plant respondent to clones pgCit009, pgCit046, and pgCit048, were determined. These segregation data were analysed together with all other markers. Two of these clones could be mapped at a 0.001 significance level, with pgCit048 linked to the isozyme locus *sdh* within 20 cM, and pgCit009 linked to a large linkage group (see linkage map in Fig. 5-3). The locus of pgCit046 showed linkage only at the 0.01 significance level, and therefore was not used in map construction, based on the reason addressed in the next section. As mentioned in the last section, the two mapped clones were also used in another mapping study with a different population derived from *Citrus* crossed with citrumelo, where these two clones followed Mendelian inheritance (R. Durham, unpublished data). A comparison between these two maps was made. It was found, in the citrumelo-derived map, that clone pgCit048 was also linked to the SDH isozyme locus with 5 cM distance between them, and pgCit009 was linked to a large linkage group, also. Much similarity was found between the two large pgCit009-linked groups of these two maps. The cumulative results lend strong support to the transposon hypothesis suggested in the last section.

Map Construction

When the linkage analysis results provided by LINKAGE-1 were used to construct a linkage map, it was found that only those linked loci determined by a 0.001 significance level ($P < 0.001$) were useful for determining gene orders to make map construction possible. Those linked loci determined by 0.05 or 0.01 significance levels did not follow the principle of three-point-cross. Therefore the 0.001 significance level was used as the threshold to determine linkage. Using a larger population, Torres et al. (1985) reported the same situation in their isozyme linkage study, but the threshold suitable for determining linkage in their study was determined to be 0.05 significance level. The difference of the threshold might have resulted from the difference in population sizes, because the larger population will give a better inferrence in statistical analysis.

Figure 5-3 shows the map constructed by this approach. Eight linkage groups including 35 loci were mapped, and 7 loci were not linked in this study. There were two markers, pgCit009 and pgCit048, that were mapped based on the transposon hypothesis discussed in the previous two sections of this chapter. The mapped loci comprised 29 RFLP markers and all of the 6 isozyme markers used in this study. The mapped RFLP loci comprised 6 cDNA markers and 23 genomic DNA markers. For convenience, the 8 linkage groups were temporarily numbered as 1-8 from the largest to the smallest (Fig. 5-3). The average length of linkage group was 39.2 cM. The largest linkage group comprised 12 loci and covered 118 cM; 4 linkage groups comprised 2 loci in each and the smallest one covered 5 cM. The whole constructed map covered 314 cM or 474 cM without/with 10 cM borders on each end of the linkage group. The average distance between two adjacent loci

Fig. 5-3: *Citrus* linkage map comprising RFLP and isozyme markers. Asterisks indicate loci that deviated from the expected monogenic segregation ratios. Clones pRLc32 and pgCit016 lack sufficient information, therefore could not be located. Clone pRLc32 was linked to pRLc27 ($P < 0.001$) and pgCit010 ($P < 0.01$), but without evidence of linkage to other loci. Clone pgCit016 was linked to pgCit028, pgCit021, and pgCit009 with P value < 0.001, linked to pgCit069 with P value < 0.01, and linked to *pgm-1* and pgCit010 with P value < 0.05; but without good fitness for the use of three-point-cross method.


1
pRLc32*
pRLc89*
4.5
pgCit062*
10.3
pgCit053
8.9
pRLc27*
21.2
pgCit010
6.3
pgm-1
21
pgCit069
32.7
? (pgCit016)
pgCit028
7
pgCit021
6.3
pgCit009
2
pgCit035
19
pgCit017
10
pgCit049
7.7
pgCit041
12.3
pgCit026
3
pgCit001L
30
pcPt001
3.3
acp
12
pgCit037
4
pgCit020*
13.3
pRLc24*
4.8
pgCit056*
idh
9
pgCit025*
7.7
pgCit015*
5
got-1
28
pgCit074
6
pgCit054S
26.5
pgCit005
8
pgCit048
5
sdh
7
pgi
7.3
pRLc91
Unlinked: pgCit011, 019*, 027, 031S, 046, 054L, 071*

was 11.6 cM. The largest distance between two adjacent loci was 32.7 cM (linkage group 1, Fig. 5-3), and the smallest was 0 cM (group 4).

Of the 11 loci with distorted segregation (Table 5-1), 9 were mapped in this study. It is of interest to note that these distorted segregation loci were clustered in two areas of the map (Fig. 5-3). There were 4 distorted segregation loci located near the pRLc32 end of linkage group 1. In linkage group 4, all of the 5 RFLP loci showed distorted segregation, and the one non-RFLP loci, i.e. *idh* isozyme locus, segregated in its expected ratio (1:1) with a P value of 0.0548 (significance level, Table 5-2). Statistically, such a P value indicates distorted segregation, because it is only slightly greater than the theshold of significance used in this study. Although the reason for segregation distortion is unknown, it might indicate that within some areas of the *Citrus* genome, recombination is somewhat disturbed.

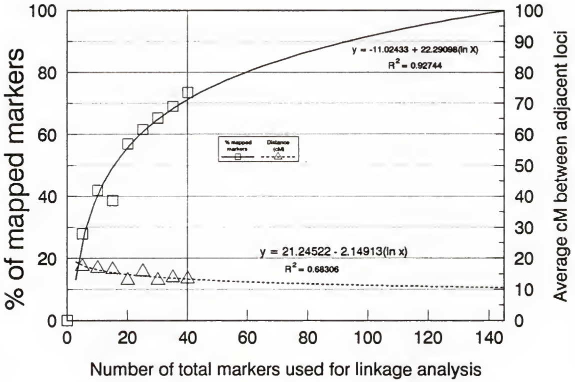
Estimation of Genome Size From Partial Linkage Maps

In the linkage analysis, the percentage of mapped markers increased with the increase of total markers, while the average distances between the adjacent loci in the maps decreased. At genome saturation, 100% of the markers added for linkage analysis can be mapped, because for every location in the genome there will be at least one mapped locus close enough to be detected by linkage analysis. It was thought useful to study the tendency of increasing percentages of mapped markers to predict the lowest marker number that would allow 100% of markers to be mapped, and the average distance between loci at that number. This approach was believed to be useful for determining the functions $P(N)$ and $D(N)$ (see Materials and Methods), which should allow the derivation of a rough estimate of the total genetic length (genome size) of *Citrus*.

Eight sets of random data containing 5, 10, 15, 20, 25, 30, 35, and 40 markers, respectively, with each set having 5 replicates, were derived from the 42 available random markers listed in Tables 5-1 and 5-2, and were used to estimate the genome size. These 40 (8 x 5) data groups were analysed separately with LINKAGE-1. The percentage of mapped markers, number of linkage groups, average distance of adjacent loci, and coverage of the map (in cM) in each case were calculated. The average percentages of mapped markers and the average distances between adjacent loci in each group were regressed on their group sizes, respectively. Four different regression types, including linear, exponential, logarithmic, and power regression, were performed with a Lotus Freelance Plus program. The equations with the highest R^2 values were chosen as the best functions of $P(N)$ and $D(N)$.

The best candidate equations for the functions $P(N)$ and $D(N)$ were determined as $y = -11.02433 + 22.29098(\ln x)$ and $y = 21.24522 - 2.14913(\ln x)$, respectively (Fig. 5-4). Both were derived from logarithmic regressions and had R^2 values of 0.92744 and 0.68306 respectively. If each random marker set being totally mapped, i.e. $P(N) = 100\%$, indicates that the saturation of genome (by the mapped loci) has been reached, then the "x" value that makes the "y" value 100% would be the lowest number of markers needed to saturate the genome. According to the first equation ($y = -11.02433 + 22.29098\ln x$), when $x = 145.57$, $y=100$; this implies $P(N) = 100\%$, when $N = 145.57$. Thus the lowest number for saturating the genome was estimated to be ca. 145. At this number, the distance between adjacent loci was estmated to be ca. 10.5 cM, because $D(N) = 10.54$, when $N = 145.57$. Therefore, the *Citrus* genome size was estimated as approximately 1500 cM ($G = N \times P(N) \times D(N) = 1534.31$ cM, when $N = 145.57$). Based on this rough estimate, the percentage of the *Citrus* genome covered by the constructed linkage map (Fig.

Fig. 5-4: Plots derived from regressing the percentages of mapped markers and the average distances (between adjacent loci in the maps) on the numbers of total markers used for linkage analyses. The extension of curves are for the estimation of *Citrus* genome size.



% of mapped markers
Average cM between adjacent loci
Number of total markers used for linkage analysis
y = -11.02433 + 22.29098(ln X)
R² = 0.92744
y = 21.24522 - 2.14913(ln x)
R² = 0.68306
% mapped markers
Distance (cM)
0
20
40
60
80
100
120
140
10
30
50
70
90

5-3) was calculated by two different approaches. One approach was to multiply the average distance between loci by the total number of mapped loci and divide the product by the estimated genome size (i.e. 11.6 cM/locus x 35 loci /1500 cM x 100); the resulting percentage was ca. 27%. A more straightforward approach was to divide the number of mapped loci by the lowest number of loci needed for saturating the genome (i.e. 35 loci/145 loci x 100); the ensuing percentage was ca. 24%. These two different approaches gave quite similar results. The kb/cM ratio of the *Citrus* genome could be roughly estimated as ca. 375 kb/cM, because the *Citrus* genome size is 0.6 pg (Guerra, 1984), or equal to 5.63 x 10^5 kb.

Using the MAPMAKER program, Hulbert et al. (1988) estimated the genome size of *Bremia lactucae* based on a double estimation; the distance between the linked loci for inferring the LOD score values was used first; and the expected numbers of pairs of markers in the genome at specific LOD score values were used second. With this approach, using the data derived from a partial linkage map that spanned over 230 cM (i.e. 11.5% of the estimated genome size), the *Bremia lactucae* genome size was estimated to be ca. 2000 cM. The estimation of the *Citrus* genome size performed here also largely depended on two factors. One was the estimate of loci number for saturating the genome, and the other was the estimate of the distance between the adjacent loci. In such a case, a deficiency of reliability of either factor will reduce the precision of the estimate. From this point of view, the estimate of the *Citrus* genome size and any other data derived from this estimate needs to be used with care, because of the relatively lower R^2 value in the equation determined for the function *D(N)*, and the somewhat long extrapolation of the x value in the equations for making the estimates.

CHAPTER 6

SUMMARY AND CONCLUSION

A Simplified Method for Enhancing RFLP Studies in Plants

The protocols developed in the course of this research (Chapter 3) may be of value for others initiating a similar project. The first step was a simplified procedure that can be used for creating a genomic library that is low in organellar clones and enriched for single copy genes. There were three special features of this protocol: 1) treatment with 50% ethanol followed by Triton lysis for eliminating most of the organellar DNA, 2) use of two CsCl gradients for DNA purification, and 3) *Pst*I-digested genomic DNA fragments for repetitive sequence elimination during library construction. After library construction, the Short Miniprep (SMP) was used to make the selection of appropriate probes for plant RFLP studies more efficient, by quickly screening colonies for those harboring inserts larger than 500 bp. This method allowed a quick overall survey of the library immediately after its construction. It has been estimated that a working team of two persons could easily examine up to 500 colonies in one week using this protocol.

Citrus Genome Organization

The *Citrus* genome size has been estimated to be 0.6 pg/C of DNA or 5.63 x 10^5 kb (Guerra, 1984). The size of the *Citrus* genome was roughly estimated, according to the recombination frequency approach described in Chapter 5, to be ca.

1500 cM. The relationship between physical and genetic distances, i.e. the average kb/cM ratio, could be roughly estimated as ca. 375 kb/cM.

The *Citrus* genome is comprised of methylated and nonmethylated portions. The non-methylated portion was estimated to be comprised of 73% single copy sequences, 18.9% multiple copy sequences, and 8.1% repetitive sequences, using the high stringency wash in Southern analysis. Some clones revealed more copy numbers in moderate stringency, rather than the high stringency washes. These results suggest that some *Citrus* genes have evolved into multigene families.

The Possible Role of Transposons in Citrus RFLPs and Genome Evolutionary Dynamics

The fact that many traits in some *Citrus* species are genetically unstable might imply special genetic characteristics. Examples of instability given by classical genetic studies include the high frequency of bud variations that may occur (Hodgson, 1967; Soost and Cameron, 1975), re-occurring mutations in different cases (Shamel, 1943; Hodgson, 1967), and the high frequencies of commonly found chimeras (Soost and Cameron, 1975; Bowman et al., 1989).

The features of some RFLPs found in *Citrus* strongly suggested that insertion/deletion is an active mechanism in the genome. At least 30% of RFLPs were caused by such a mechanism. A new model was proposed in Chapter 4 to explain banding-pattern changes observed in the insertion/deletion events; this model may be useful for examining the insertion/deletion phenomena in other species, also. There were 3 clones that showed extra alleles in progeny analyses in this study, apparently violating Mendelian genetics and preventing recombination-frequency mapping. A progeny analysis of a genomic clone, pgCit075L, implied the

existence of a sequence outside the nucleus that was homologous to a repetitive sequence in the nuclear genome.

All of these puzzling phenomena that were inexplicable with general genetic principles could be explained with the transposon assumption. Perhaps the best explanation for clone pgCito57L was that this sequence also exists in the mitochondria in the form of an episome. If true, it may actually be a transposon-like sequence (Kemble et al. 1983; Nargang et al. 1984; Kadowaki et al. 1988). With the assumption of transposition, the "normal" genotypes of the 3 clones showing extra alleles in progeny analyses could be determined, and their segregations were just as expected according to chi-square tests. Moreover, 2 of these 3 markers were mapped in this study, and the mapping results were consistent with the other mapping study using different plant populations where these clones followed Mendelian genetics (R. Durham, personal communication). The well known genetic instability of *Citrus* might be related to the activity of transposon(s); likewise, transposon(s) might have been an important factor in *Citrus* genome evolutionary dynamics.

Genetic Map of Citrus

Little knowledge about *Citrus* genetics has been provided through classic genetic studies. Breeding efforts in *Citrus* have long been hindered by factors as reviewed in Chapter 2 of this study. The construction of a linkage map in *Citrus* based on RFLPs offers new opportunities for applications in genetics and breeding. With further efforts, the *Citrus* genome can be saturated with RFLP markers, and genes governing horticulturally important traits can be mapped based on linkage with RFLP markers. With the availability of such a map in *Citrus* and the vast

collection of germplasm, in conjunction with the fast advances in gene manipulation, the efficiency of *Citrus* variety improvement can be greatly enhanced in the future.

APPENDIX

Probe selection for RFLP studies.

Clone No.	kb	EcoRI	BglII	EcoRV	HindIII	BamHI	Pst I	copy[1] no.
		Enzyme						
A. Clones only tested with high stringency wash (0.1 x SSC).								
pgCit001L	1.6	-[2]	-	+[3]				S
2	3.0	-	?[4]	?				S
3	1.0	-	-	-	-	-	-	M
4	0.4	-	?	-				S
5	1.8	+	+	+	-	-	-	S
6	3.2	-	-	-	?	-	?	M
7	3.0	?	?	?				S
8	0.7	-	-	-	-	-	-	M
9	1.8	+	-	-	+	-	-	S
10	2.25	+	+	+	+	?	+	S
11	3.5	-	-	+	-	+	?	S
12	5.0	-	-	-	-	?	-	S
13	4.0	-	-	-	-	-	-	S
14	2.2	-	-	-	-	-	-	S
15	1.5	+	+	+	+	+	?	S
16	2.2	+	?	+	+	+	-	M
17	1.55	+	-	+				S
18	1.3	+	+	+				S
21	2.0	+	-	+				S
22	1.9	-	-	-	+	-	-	S
23	1.9	-	-	-	-	-	-	R
24	4.5	-	-	-				R
25	2.1	-	-	+	+	-	+	S
26	2.25	+	+	-	+	+	?	S
27	2.2	+	+	+				S
28	2.15	+	+	-	-	?	-	S
29	4.0	-	-	-				S
30	1.8	-	+	+	-	+	-	S
31L	3.3	-	-	-	-	-	-	R
31S	2.5	-	+	-	-	-	-	S
32	1.8	?	?	+	+	+	-	M
33	1.1	-	-	-				S
34	3.5	-	-	-				R

35	1.3	-	+	-				S
36	1.4	-	-	-				S
37	2.3	+	+	+	+	?	+	S
38	3.1	-	?	-	+	-	-	M
39	2.0	+	-	-	?	-	-	S
45	1.6	+	-	-				M
48	4.0	+	-	-				S
49	2.0	+	+	-				S
51	4.0	-	+	+				S
52	1.0	-	+	-				S
57L	3.2	-	+	+				S
58	0.6	-	?	?				S
59	0.5	-	-	-				S
60	3.2	-	-	-				S
61	2.5	+	+	+	+	+	+	M
62	2.3	+	?	-	?	-	-	M
63	1.3	+	?	-	-	?	?	S
64	1.9	-	-	-	+	-	-	S
65	2.0	-	-	-	-	-	-	M
66	4.0	-	?	?				R
67	2.0	-	-	-				M
68	5.0	-	-	+				S
69	2.1	?	+	+	+	+	?	S
70	1.7	-	-	-	-	-	-	S
71	3.2	+	-	+				S
72L	4.2	-	+	-				S
73	>10.0	-	-	-	-	-	-	R
74	1.3	+	?	?	?	?	?	S
75L	1.2	-	-	-				R
pcPt001		+	+	-				S
002		-	-	-	-	-	-	S
pRLc24	0.95	-	+	-	-	-	?	S
25	0.8	-	-	-				S
27L	0.74	-	-	-				S
31	0.62	-	-	-	-	-	-	S
32	1.23	-	+	-				M
39	1.35	-	+	+				M
60	0.55	-	-	-	-	-	-	S
91	0.74	-	-	-				S
94	1.23	-	-	+	-	+	-	M

B. Clones tested with moderate (1 x SSC) and high stringency washes.

pgCit 19	1.35	+	+	-				S
(19)[5]		+	-	-				M
20	1.0	-	-	-				S
(20)		+	+	+				M
40	0.5	+	?	-				M
(40)		+	-	?				M
41	0.2	-	+	+	+	-	-	S

(41)		-	+	+	+	-	-	S
42	0.5	-	-	-				S
(42)		-	-	-				S
43	2.0	-	-	-				M
(43)		-	-	-				M
44	2.1	-	?	?				M
(44)		-	+	+				M
46	4.3	-	-	+				S
(46)		-	-	+				S
47	1.0	-	-	-				M
(47)		-	-	-				R
50	0.8	-	-	-				S
(50)		-	-	-				S
53	2.0	-	-	-				S
(53)		-	-	-				S
54L	1.2	+	-	+				S
(54L)		+	-	+				S
54S	1.0	+	-	+				S
(54S)		+	-	+				S
55	0.8	-	-	-				S
(55)		-	-	-				S
56	2.0	+	+	+				S
(56)		+	+	+				S
pRLc 10	0.18	-	-	-				R
(10)		+	-	-				R
11	0.5	-	+	-	-	-	-	R
(11)		-	-	-	-	-	-	R
38	0.35	-	-	-				S
(38)		-	-	-				S
40	1.4	-	-	-				S
(40)		-	-	-				S
41	0.4	-	-	-				M
(41)		-	+	-				M
66L	0.8	-	-	-				S
(66L)		-	-	-				S
66S	0.25	-	-	-				S
(66S)		-	-	-				S
89	0.4	-	+	+				M
(89)		-	+	+				M
90	0.375	-	-	-				S
(90)		-	-	-				R

[1] S: single copy; M: multiple copy; R: repetitive sequence.

[2] Not polymorphic between Clementine and LB1-21 hybrid.

[3] Polymorphic between Clementine and LB1-21 hybrid.

[4] Not conclusive for polymorphism.

[5] Washed with moderate stringency condition.

LITERATURE CITED

Adams, W.T. 1983. Application of isozymes in tree breeding. In: S.D. Tanksley and T.J. Orton (eds.). Isozymes in plant genetics and breeding, Part A. Elsevier. New York. pp. 381-400.

Allard, R.W. 1956. Formulae and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235-278.

Arulsekar, S. and D.E. Parfitt. 1986. Isozyme analysis procedures for stone fruit, almond, grape, walnut, pistachio, and fig. HortScience 21(4):928-933.

Arulsekar, S., D.E. Parfitt., W. Beres, and P.E. Hansche. 1986. Genetics of malate dehydrogenase isozymes in the peach. J. Hered. 77:49-51.

Avise, J.C., R.A. Lansman, and R.O. Shade. 1979. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. Genetics 92:279-295.

Barker, D., E. Wright, K. Nguyen, L. Cannon, P. Fain, D.Goldgar, D.T. Bishop, J. Carey, B. Baty, J. Kivlin, H. Willard, J.S. Waye, G. Greig, L. Leinward, Y. Nakamura, P. O'Connell, M. Leppert, J.-M. Lalouel, R. White, and M. Skolnick. 1987. Gene for von Recklinghausen neurofibromatosis is in the pericentromeric region of chromosome 17. Science 236:1100-1102.

Beckmann, J.S. and M. Soller. 1986. Restriction fragment length polymorphisms and genetic improvement of agricultural species. Euphytica 35:111-124.

Beltz, G.A., K.A. Jacobs., T.H. Eickbush, P.T. Cherbas, and F.C. Kafatos. 1983. Isolation of multigene families and determination of homologies by filter hybridization methods. Methods in Enzymology 100:266-285.

Ben-Hayim, G., A. Shani, and A. Vardi. 1982. Evaluation of isozyme systems in *Citrus* to facilitate identification of fusion products. Theor. Appl. Genet. 64:1-5.

Bernatzky, R. and S.D. Tanksley. 1986. Majority of random cDNA clones correspond to single loci in the tomato genome. Mol. Gen. Genet. 203:8-14.

Bolkan, H. A., V.M. Williamson, and C.M. Waters.1987. Use of cellulose acetate electrophoresis as an alternative to starch gel electrophoresis for detecting root-knot nematode resistance in tomato. Plant Disease 71(11):1001-1003.

Bonierbale, M.W., R.L. Plaisted, and S.D. Tanksley. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095-1103.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314-331.

Bowman, K.D., F.G. Gmitter, Jr, and G.A. Moore. 1989. Natural fruit chimeras as a resource for Citrus cultivar development (abstract). ASHS Annual Meeting/ Program and Abstracts. p. 138.

Brewer, G.J., C.F. Sing, and E.R. Sears. 1969. Studies of isozyme patterns in nullisomic-tetrasomic combinations of hexaploid wheat. Proc. Natl. Acad. Sci. 64:1224-1229.

Burke, J. H. 1967. The commercial *Citrus* regions of the world. In: W. Reuther, H.J. Webber, and L.D. Batchelor (eds.). The citrus industry, Vol. 1. Univ. of California. Berkeley, Calif. pp. 40-189.

Burr, B., F.A. Burr, K.H. Thompson, M.C. Albertson, and C.W. Stuber. 1988. Gene mapping with recombination inbreds in maize. Genetics 118:519-526.

Burr, B., S. V. Evola, and F. A. Burr. 1983. The application of restriction fragment length polymorphism to plant breeding. In: J.K. Stelow and A. Hollaender (eds.). Genetic engineering: Principles and methods, Vol. 5. Plenum Press. New York and London. pp. 45-59.

Cabera, A. M. and H. Lima. 1981. Peroxidase activity during fruit development in the so-called "Acidic fruit" of citrus. Proc. Int. Soc. Citriculture 1:13-17.

Cameron, J.W. and H.B. Frost. 1968. Genetics, breeding, and nucellar embryony. In: W. Reuther, L.D. Batchelor, and H.J. Webber (eds.). The citrus industry, Vol. 2. Univ. of California. Berkeley, Calif. pp. 325-370.

Cameron, J.W. and R.K. Soost. 1984. Citrus. In: N.W. Simonds (ed.). Evolution of crop plants. Longman. New York. pp. 261-264.

Chang, C., J.L. Bowman, A.W. DeJohn, and E.S. Lander. 1988. Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 85:6856-6860.

Chaparro, J.X., R.E. Durham., G.A. Moore, and W.B. Sherman.1987. Use of isozyme techniques to identify peach x "nonpareil" almond hybrids. HortScience 22(2):300-302.

Chase, C.D. and D.R. Pring. 1985. Circular plasmid DNAs from mitochondria of *Sorghum bicolor*. Plant Mol. Biol. 5:303-311.

Chase, C.D. and D.R. Pring. 1986. Properties of the linear N1 and N2 plasmid-like DNAs from mitochondria of cytoplasmic male-sterile *Sorghum bicolor*. Plant Mol. Biol. 6:53-64.

Christiansen, C. 1984. Molecular cloning in bacteriophage lambda and in cosmids. In: J.M. Walker (ed.), Methods in molecular biology, Vol. 2. Nucleic acid . Humana Press. Clifton, New Jersey. pp. 313-319.

Church, G. M. and W. Gilbert. 1984. Genomic sequencing. Proc. Natl. Acad. Sci. 81:1991-1995.

Cohen, S.N., A.C.Y. Chang, and L.Hsu. 1973. Non-chromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. 69:2110-2114.

Craddock, E.M. and W.E. Johnson. 1976. First gene mapping in Hawaiian Drosophila. The American Naturalist 110(975):861-865.

Crouse , J. and D. Amorese. 1987. Ethanol precipitation: ammonium acetate as an alternative to sodium acetate. Focus 9(2):3-5.

Davis, L.G., M.D. Dibner, and J.F. Battey. 1986. Basic methods in molecular biology. Elsevier Sci. Pub. New York. 372 pp.

Dellaporta, S. L., J. Wood, and J. B. Hicks. 1983. A plant DNA minipreparation: Version II. Pl. Mol. Biol. Rep. 1(4):19-21.

Donis-Keller, H., P. Green, C. Helms, S. Cartinhour, B. Weiffenbach, K. Stephens, T.P. Keith, D.W. Bowden, D.R. Smith, E.S. Lander, D. Botstein, G. Akots, K.S. Rediker, T. Gravius V.A. Brown, M.B. Rising, C. Parker, J.A. Powers, D.E. Watt, E.R. Kauffman, A. Bricker, P. Phipps, H. Muller-Kahle, T.R. Fulton, S. Ng, J.W. Schumm, J.C. Braman, R.G. Knowlton, D.F. Barker, S.M. Crooks, S. E. Lincoln, M.J. Daly, and J. Abrahamson. 1987. A genetic linkage map of the human genome. Cell 51:319-337.

Doyle, J.J., J.L. Doyle, and L.H.B. Hortorium. 1990. Isolation of plant DNA from fresh tissue. Focus 12(1):13-15.

Durham, R.E., G.A. Moore, and W.B. Sherman. 1987. Isozyme banding patterns and their usefulness as genetic markers in peach. J. Amer. Soc. Hort. Sci. 112(6):1013-1018.

Ellis, T.H.N. and W. Cleary. 1988. Trisomy: A useful adjunct to RFLP mapping in pea. Chromosoma 96:91-94.

Elston, R.C. and J. Stewart. 1971. A general model for the analysis of pedigree data. Hum. Hered. 21:523-542.

Esen, A. and R.W. Scora. 1975. Distribution of enzymatic browning of young shoot homogenates in the Aurantioideae. Amer. J. Bot. 62(10):1078-1083.

Esen, A., R.W. Scora, and R.K. Soost. 1975. A simple and rapid screening procedure for identification of zygotic *Citrus* seedlings among crosses of certain taxa. J. Amer. Soc. Hort. Sci. 100(5):558-561.

Esen, A. and R.K. Soost. 1972. Aneuploidy in citrus. Amer. J. Bot. 59:473-477.

Esen, A. and R.K. Soost. 1974a. Inheritance of browning of young-shoot extract of *Citrus*. J. Hered. 65:97-100.

Esen, A. and R.K. Soost. 1974b. Polyphenol oxidase-catalyzed browning of young shoot extracts of citrus taxa. J. Amer. Soc. Hort. Sci. 99(6):484-489.

Esen, A. and R.K. Soost. 1977. Separation of nucellar and zygotic citrus seedlings by use of polyphenol oxidase-catalyzed browning. Proc. Int. Soc. Citriculture 2:616-618.

Fedoroff, N.V. 1983. Controlling elements in maize. In: J.A. Shapiro (ed.). Mobile genetic elements. Academic Press, Inc. New York. pp. 1-58.

Fedoroff, N.V. 1984. Transposable genetic elements in maize. Sci. Amer. 250(6):84-98.

Fisher, R.A. and B. Balmukand. 1928. The estimation of linkage from the offspring of selfed heterozygotes. J. Genet. 20(1):79-92.

Fobes, J.F. 1980. Trisomic analysis of isozymic loci in tomato species: segregation and dosage effects. Biochem. Genet. 18:401-421.

Frost, H.B. and R.K. Soost. 1968. Seed reproduction: Development of gametes and embryos. In: W. Reuther, L.D. Batchelor, and H.J. Webber (eds.). The citrus industry, Vol. 2. Univ. of California. Berkeley, Calif. pp. 290-324.

Galletta, G.J. 1983. Pollen and seed management. In: J.N. Moore and J. Janick (eds.). Methods in fruit breeding. Purdue Univ. West Lafayette, Indiana, pp. 23-47.

Geraci, G., L.A. Manzocchi, N. Tusa, G. Occorso, L. Radogna, and F. De Pasquale. 1981. Comparison of different methods for identifying zygotic and nucellar seedling in citrus. Proc. Int. Soc. Citriculture 1:1-4.

Gergen, J.R., R.H. Stern, and P.C. Wensink. 1979. Filter replicas and permanent collections of recombinant DNA plasmids. Nucleic Acids Res. 7:2115-2136.

Gilliam, T.C., M. Bucan, M.E. MacDonald, M. Zimmer, J.L. Haines, S.V. Cheng, T.M. Pohl, R.H. Meyers, W.L. Whaley, B.A. Allitto, A. Faryniarz, J.J. Wasmuth, A.-M. Frischauf, P.M. Conneally, H. Lehrach, and J.F. Gusella. 1987. A DNA segment encoding two genes very tightly linked to Huntington's disease. Science 238:950-952.

Gmitter, F.G. 1985. *Citrus* embryogenesis *in vitro*: Culture establishment, plant regeneration, and phenotypic characterization. Doctorial dissertation. University of Florida. 142 pp.

Goblet, J-P., M. Boutry, G. Duc, And M. Briquet. 1983. Mitochondrial plasmid-like molecules in fertile and male sterile *Vicia faba* L. Plant Mol. Biol. 2:305-309.

Goodenough, U. 1984. Genetics. 3rd ed. Saunders College Pub. New York. 894 PP.

Gottlieb, L.D. 1982. Conservation and duplication of isozymes in plants. Science 216(23):373-380.

Gottlieb, L.D. 1983. Isozyme number and phylogeny. In: U. Jensen and D.E. Fairbrothers (eds.). Proteins and nucleic acids in plant systematics. Springer-Verlag, Berlin. pp. 209-221.

Grierson, D. and S. N. Covey. 1984. Plant molecular biology. Blackie. Glasgow and London. 176 pp.

Grosser, J.W., F.G. Gmitter, Jr, and J.L. Chandler 1988a. Intergeneric somatic hybrid plants of *Citrus sinensis* cv. Hamlin and *Poncirus trifoliata* cv. Flying Dragon. Plant Cell Reports 7:5-8.

Grosser, J.W., F.G. Gmitter, Jr. and J.L. Chandler 1988b. Intergeneric somatic hybrid plants from sexually incompatible woody species: *Citrus sinensis* and *Severinia disticha*. Theor. Appl. Genet. 75:397-401.

Guerra, M. dos S. 1984. Cytogenetics of Rutaceae. II. Nuclear DNA content. Caryologia 37(3):219-226.

Hake, S., E. Vollbrecht, and M. Freeling. 1989. Cloning Knotted, the dominant morphological mutant in maize using Ds2 as a transposon tag. EMBO J. 8(1):15-22.

Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166:557-580.

Hearn, C.J. 1973. Development of scion cultivars of citrus in Florida. Proc. Fla. State Hort. Soc. 86:84-88.

Hearn, C.J. 1977. Recognition of zygotic seedlings in certain orange crosses by vegetative characters. Proc. Int. Soc. Citriculture 2:611-614.

Hearn, C.J. 1986. Development of seedless grapefruit cultivars through budwood irradiation. J. Amer. Soc. Hort. Sci. 111(2):304-306.

Helentjaris, T. 1987. A genetic linkage map for maize based on RFLPs. Trends in Genetics 3(8):217-221.

Helentjaris, T., G. King, M. Slocum, C. Siedenstrang, and S. Wegman. 1985. Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Pl. Mol. Biol. 5:109-118.

Helentjaris, T., M. Slocum, S. Wright, A. Schaefer, and J. Nienhuis. 1986. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphismns. Theor. Appl. Genet. 72:761-769.

Hirai, M. and I. Kozaki. 1981. Isozymes of citrus leaves. Proc. Int. Soc. Citriculture 1:10-13.

Hodgson, R.W. 1967. Horticultural varieties of citrus. In: W. Reuther, H.J. Webber, and L.D. Batchelor (eds.). The citrus industry, Vol. 1. Univ. of California. Berkeley, Calif. pp. 431-591.

Holl, F.B. 1975. DNA isolation from plants for use in DNA feeding experiments. In: O.L. Gamborg, and L.R. Wetter (eds.). Plant tissue culture methods. Natl. Res. Council Canada. Saskatoon, Saskatchewan. pp.65-69.

Huang, F.H., S. Tsai, and R.C. Rom. 1984. An electrophoresis method for water-soluble protein of Prunus. HortScience 19(2):242-243.

Hudson, R.R. 1982. Estimating genetic variability with restriction endonucleases. Genetics 100:711-719.

Hudson, R.R. 1989. How often are polymorphic restriction sites due to a single mutation? Theor. Popul. Bio. 36:23-33.

Hulbert, S.H., T.W. Ilott, E.J. Legg, S.E. Lincoln, E.S. Lander, and R.W. Michelmore. 1988. Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphisms. Genetics 120:947-958.

Iglesias, L., H. Lima, and J.P. Simon. 1974. Isoenzyme identification of zygotic and nucellar seedlings in citrus. J. Hered. 65:81-84.

Immer, F.R. 1930. Formulae and tables for calculating linkage intensities. Genetics 15:81-98.

Immer, F.R. and M.T. Henderson. 1943. Linkage studies in barley. Genetics 28:419-440.

Jones, I.M., S.B. Primrose, A. Robinson, and D.C. Ellwood. 1981. Effect of growth rate and nutrient limitation on the transformability of *Escherichia coli* with plasmid deoxyribonucleic acid. J. Bacteriology 146(3):841-846.

Kadowaki, K., K. Yazaki, T. Osumi, K. Harada, M. Katsuta, and M. Nakagahra. 1988. Distribution of mitochondrial plasmid-like DNA in cultivated rice (*Oryza sativa* L.) and its relationship with varietal groups. Theor. Appl. Genet. 76:809-814.

Kam-Morgan, L.N.W. and B.S. Gill. 1989. DNA restriction fragment length ploymorphisms: A strategy for genetic mapping of D genome of wheat. Genome 32:724-732.

Kaneko, T., T. Terachi, and K. Tsunewaki. 1986. Studies on the origin of crop species by restriction endonuclease analysis of organellar DNA. II. Restriction analysis of ctDNA of 11 *Prunus* species. Jpn. J. Genet. 61:157-168.

Keim, P. 1987. Isolation of nuclei from soybean suspension cultures. Methods in Enzymology 148:535-541.

Kemble, R.J. and J.R. Bedbrook. 1980. Low molecular weight circular and linear DNA in mitochondria from normal and male-sterile *Zea mays* cytoplasm. Nature 284:565-566.

Kemble, R.J., R.J. Mans, S. Gabay-Laughnan, and J.R. Laughnan. 1983. Sequences homologous to episomal mitochondrial DNAs in the maize nuclear genome. Nature 304(25):744-747.

Kislev, N. and I. Rubenstein. 1980. Utility of ethidium bromide in the extraction from whole plants of high molecular weight maize DNA. Plant Physiol. 66:1140-1143.

Klug, W.S. and M.R. Cumming. 1986. Concepts of genetics. 2nd ed. Scott, Foresman, and Company. Glenview, Illinois. 699 pp.

Kobayashi, S. and H. Uchimiya. 1989. Expression and integration of a foreign gene in orange (*Citrus sinensis* Osb.) protoplasts by direct DNA transfer. Jpn. J. Genet. 64:91-97.

Kramer, H.H. and C.R. Burnham. 1947. Methods of combining linkage intensity values from backcross, F2 and F3 genetic data. Genetics 32:379-390.

Lander, E.S. and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Landry, B.S., R.V. Kesseli, B. Farrara, and R.W. Michelmore. 1987. Genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331-337.

Landry, B. S. and R. W. Michelmore. 1985. Selection of probes for restriction fragment length analysis from plant genomic clones. Pl. Mol. Biol. Rep. 3(4):174-179.

Lathrop, G. M. and J.M. Lalouel. 1984. Easy calculations of lod scores and genetic risks on small computer. Am. J. Hum. Genet. 36:460-465.

Lathrop, G. M., J.M. Lalouel, C. Julier, and J. Ott. 1985. Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am. J. Hum. Genet. 37:482-498.

Lathrop, G.M., J.M. Lalouel, and R.L. White. 1986. Construction of human linkage maps: Likelihood calculations for multilocus linkage analysis. Genet. Epidemiol. 3:39-52.

Levin, D.A. and S.W. Funderburg. 1979. Genome size in angiosperms: Temperate versus tropical species. The American Naturalist 114(6):784-795.

Li, W., S. He, and G. Liu. 1987. A study of citrus natural resources in Hunan by analyses of leaf isozymes (in Chinese). Acta Hort. Sinica 14(3):153-160.

Lichter, P., C-J.C. Tang, K. Call, G. Hermanson, G.A. Evans, D. Housman, and D.C. Ward. 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. Science 247:64-69.

Luthe, D.S. and R.S. Quatrano. 1980. Transcription in isolated wheat nuclei. I. Isolation of nuclei and elimination of endogenous ribonuclease activity. Plant Physiol. 65:305-308.

Mandel, M. and A. Higa. 1970. Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53:159-162.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York. 545 pp.

Mather, K. 1938. The measurement of linkage in heredity. Methuen & Co., Ltd. London. 132 pp.

Mathew, C.G.P. 1984. The isolation of high molecular weight eukaryotic DNA. In: J.M. Walker (ed.). Methods in molecular biology, Vol. 2, Nucleic acids. Humana Press. Clifton, New Jersey. pp. 31-34.

McCouch, S.R., G. Kochert, Z.H. Yu, Z.Y. Wang, G.S. Khush, W.R. Coffman, and S.D. Tanksley. 1988. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76:815-829.

McMillin, D.E. 1983. Plant isozymes: a historical perspective. In: S.D. Tanksley and T.J. Orton (eds.). Isozyme in plant genetics and breeding, Part A. Elsevier. New York. pp. 3-13.

McMillin, D.E., D.G. Roupakias, and J.G. Scandalios. 1979. Chromosomal location of two mitochondria malate dehydrogenase structural genes in *Zea mays* using trisomics and B-A translocations. Genetics 92:1241-1250.

Mederic, C., O. Bertaux, J.D. Rouzeau, and R. Valencia. 1987. Isolation of high molecular weight DNA from whole *Euglena* cells (English abstract, In: Biosis/cas selects: Plant Genetics Issue ll, 1988 P.24). Protoplasma 141(2/3):139-148

(in French).

Medgyesy, P., E. Fejes, and P. Maliga. 1985. Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. Proc. Natl. Acad. Sci. USA. 82:6960-6964.

Mettler, I.J. 1987. A simple and rapid method for minipreparation of DNA from tissue cultured plant cells. Pl. Mol. Biol. Rep. 5(3):346-349.

Meyerowitz, E.M. and R.E. Pruitt. 1985. *Arabidopsis thalinana* and plant molecular genetics. Science 229:1214-1218.

Moore, G.A. and G.B. Collins. 1983. New challenges confronting plant breeders. In: S.D. Tanksley and T.J. Orton (eds.). Isozymes in plant genetics and breeding, Part A. Elsevier. New York. pp. 25-58.

Morton, N.E., C.J. Maclean, R. Lew, and S. Yee. 1986. Multipoint linkage analysis. Amer. J. Hum. Genet. 38:868-883.

Murray, M. G. and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acid Res. 8(19):4321-4325.

Nargang, F.E., J.B. Bell, L.L. stohl, and A.M. Lambowitz. 1984. The DNA sequence and genetic organization of a *Neurospora* mitochondrial plasmid suggest a relationship to introns and mobile elements. Cell 38:441-453.

Nikiforova, I.D. and V.I. Negruk. 1983. Comparative electrophoretical analysis of plasmid-like mitochondrial DNAs in *Vicia faba* and in some other legumes. Planta 157:81-84.

Ohgawara, T., S. Kobayashi, E. Ohgawara, H. Uchimiya, S. Ishii. 1985. Somatic hybrid plants obtained by protoplast fusion between *Citrus sinensis* and *Poncirus trifoliata*. Theor. Appl. Genet. 71:1-4.

Ohri, D. and A. Kumar. 1986. Nuclear DNA amounts in some tropical hardwoods. Caryologia 39(3-4):303-307.

Ohyama, K. 1975. DNA isolation from plant protoplast. In: O.L. Gamborg, and L.R. Wetter (eds.). Plant tissue culture methods. Natl. Res. Council Canada. Saskatoon, Saskatchewan. PP. 70-73.

Ohyama, K., O.L. Gamborg, and R.A. Miller. 1972. Isolation and properties of deoxyribonucleic acid from protoplasts of cell suspension cultures of *Ammi visnaga* and carrot (*Daucus carota*). Plant Physiol. 50:319-321.

Old, R.W. and S. B. Primrose. 1985. Principles of gene manipulation: An introduction to genetic engineering, 3rd ed. Blackwell Sci. Pub. Boston, Massachusetts. 409 pp.

Oliver, S.G. and J.M. Ward. 1987. A dictionary of genetic engineering. 3rd ed. Cambridge Univ. New York. 153 pp.

Ortiz, J.M., J.L. Tadeo, J. Guerri, and J.B. Forner. 1981. Distinction between hybrid and nucellar citrus trees by analysis of their biochemical compounds. Proc. Int. Soc. Citriculture 1:4-7.

Osborn, T.C., D.C. Alexander, and J.F. Fobes. 1987. Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor. Appl. Genet. 73:350-356.

Ott, J. 1973. A computer program for linkage analysis in human pedigrees. Am. J. Hum. Genet. 25:57A.

Ott, J. 1974. Estimation of the recombination fraction in human pedigrees: Efficient computation of likelihood for human linkage studies. Am. J. Hum. Genet. 26:588-597.

Ott, J. 1976. A computer program for linkage analysis of general human pedigrees. Am. J. Hum. Genet. 28:528-529.

Palmer, J. D., C.R. Shields, D.B. Cohen, and T.J. Orton. 1983. An unusual mitochondrial DNA plasmid in the genus *Brassica*. Nature 301:725-728.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphism. Nature 335 (20):721-726.

Paul, A.-L., V. Vasil, I.K. Vasil, and R.J. Ferl. 1987. Constitutive and anaerobically induced DNase-I-hypersensitive sites in the 5' region of the maize *Adh1* gene. Proc. Natl. Acad. Sci. USA 84:799-803.

Peirce, L.C. and J.L. Brewbaker. 1973. Applications of isozyme analysis in horticultural science. HortScience 8(1):17-22.

Perbal, B. 1984. A practical guide to molecular cloning. A Wiley-Interscience Pub. New York. 554 pp.

Pieringer, A.P. and G.J. Edwards. 1964. Identification of nucellar and zygotic citrus seedlings by infrared spectroscopy. Proc. Amer. Soc. Hort. Sci. 86:226-234.

Price, H.J. 1976. Evolution of DNA content in higher plants. Bot. Rev. 42(1):27-53.

Rayburn, A. L. and B. S. Gill. 1985. Use of biotin-labeled probes to map specific DNA sequences on wheat chromosomes. J. Hered. 76:78-81.

Reed, K.C. and D.A. Mann. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. 13:7207-7221.

Rieger, R., A. Michaelis, and M. M. Green. 1976. Glossary of genetics and cytogenetics: Classical and molecular. 4th ed. Springer-Verlag. New York. 645 pp.

Rines, H.W., B.G. Gengenbach, K.L. Boylan, and K.K. Storey. 1988. DNA diversity in oat cultivars and species. Crop Sci. 28:171-176.

Rogers, S.O. and A.J. Bendich. 1988. Extraction of DNA from plant tissues. Pl. Mol. Bio. Man. A6:1-10.

Roose, M.L. 1988. Isozymes and DNA restriction fragment length polymorphisms in citrus breeding and systematics. In: R. Goren and K. Mendel (eds.). Proceedings of the sixth international citrus congress. Margraf Scientific Books. Weikersheim. pp. 57-67.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York. pp. 1.1-18.88.

Sarfatti, M., J. Katan, R. Fluhr, and D. Zamir. 1989. An RFLP marker in tomato linked to the *Fusarium oxysporum* resistance gene *I2*. Theor. Appl. Genet. 78:755-759.

Seyer, P., K.V. Kowallik, and R.G. Herrmann. 1981. A physical map of *Nicotiana tabacum* plastid DNA including the location of structural genes for ribosomal RNAs and the large subunit of ribulose bisphosphate carboxylase/oxygenase. Cur. Genet. 3:189-204.

Shamel, A.D. 1943. Bud variation and bud selection. In: H.J. Webber, and L.D. Batchelor (eds.). The citrus industry, Vol. 1. Univ. of California. Berkeley and Los Angeles, Calif. pp. 915-952.

Shapiro, J.A. and B. Cordell. 1982. Eukaryotic mobile and repeated genetic elements. Biology of the Cell 43:31-53.

Shaw, C.R. 1965. Electrophoretic variation in enzymes. Science 149:936-943.

Shaw, C.R. and R. Prasad. 1970. Starch gel electrophoresis of enzymes-a compilation of recipes. Biochemical Genetics 4:297-320.

Sherman, W.B. and P.M. Lyrene. 1983. Handling seedling populations. In: J.N. Moore and J. Janick (eds.). Methods in fruit breeding. Purdue Univ. West Lafayette, Indiana. pp. 66-73.

Slightom, J.L. and R.F. Drong. 1988. Procedures for constructing genomic clone banks. Pl. Mol. Man. Bio. Man. Kluwer Academic Publishers. Dordrecht, Belgium. A8:1-42.

Soller, M. and J.S. Beckmann. 1983. Genetic polymorphism in varietal identification and genetic improvement. Theor. Appl. Genet. 67:25-33.

Soost, R.K. and A.M. Torres. 1981. Leaf isozymes as genetic markers in citurs. Proc. Int. Soc. Citriculture 1:7-10.

Soost, R. K. and J. W. Cameron. 1975. Citrus. In: J. Janick and J.N. Moore (eds.). Advances in fruit breeding. Purdue Univ. West Lafayette, Indiana. pp. 507-540.

Soost, R.K. and T.E. Williams. 1980. Identification of nucellar and zygotic seedlings of *Citrus* with leaf isozymes. HortScience 15(6):728-729.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

Spencer, N., D.A. Hopkinson, and H. Harris. 1964. Phosphoglucomutase polymorphism in man. Nature 204:742-745.

Stevens, W.L. 1939. Tables of the recombination fraction estimated from the product ratio. J. Genet. 39:171-180.

Suiter, K.A., J.F. Wendel, and J. S. Case. 1983. Linkage-1: A PASCAL computer program for the detection and analysis of genetic linkage. J. Hered. 74:203-204.

Suzuki, D.T., A.J.F. Griffiths, and R.C. Lewontin. 1981. An introduction to genetic analysis. 2nd ed. Freeman. San Francisco. 911 pp.

Swingle, W.T. and P.C. Reece. 1967. The botany of citrus and its wild relatives. In: W. Reuther, H.J. Webber, and L.D. Batchelor (eds.). The citrus industry, Vol. 1. Univ. of California. Berkeley, Calif. pp. 190-430.

Tanksley, S.D. 1983. Gene mapping. In: S.D. Tanksley and T.J. Orton (eds.). Isozymes in plant genetics and breeding, Part A. Elsevier. Amsterdam. pp. 109-138.

Tanksley, S.D. 1984. Linkage relationships and chromosomal locations of enzyme-coding genes in pepper, *Capsicum annuum*. Chromosoma 89:352-360.

Tanksley, S.D., and J. Hewitt. 1988. Use of molecular markers in breeding for soluble solid content in tomato, a re-examination. Theor. Appl. Genet. 75:811-823.

Tanksley, S.D., H. Medina-Filho, and C.M. Rick. 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49(1):11-25.

Tanksley, S.D. and C.M. Rick. 1980. Isozymic gene linkage map of the tomato: Applications in genetics and breeding. Theor. Appl. Genet. 57:161-170.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257-264.

Torres, A.M. 1983. Fruit trees. In: S.D. Tanksley and T.J. Orton (eds.). Isozymes in plant genetics and breeding, Part B. Elsevier. Amsterdam. pp. 401-421.

Torres, A.M. and B.O. Bergh. 1980. Fruit and leaf isozymes as genetic markers in avocado. J. Amer. Soc. Hort. Sci. 105(4):614-619.

Torres, A.M., R.K. Soost, and U. Diedenhofen. 1978. Leaf isozymes as genetic markers in citrus. Amer. J. Bot. 65(8):869-881.

Torres, A.M., R.K. Soost, and T. Mau-Lastovicka. 1982. Citrus isozymes: Genetics and distinguishing nucellar from zygotic seedlings. J. Hered. 73:335-339.

Torres, A.M., T. Mau-Lastovicka., T.E. Williams, and R,K. Soost. 1985. Segregation distortion and linkage of *Citrus* and *Poncirus* isozymes genes. J. Hered. 76:289-294.

Toxopeus, H. 1962. Notes on the genetics of a few leaf characters in the genus *Citrus*. Euphytica 11(1):19-25.

Vieira, J. and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259-268.

Wahleithner, J.A. and D.R. Wolstenholme. 1987. Mitochondrial plasmid DNAs of broad bean: Nucleotide sequences, complex scondary structures, and transcription. Curr. Genet. 12:55-67.

Watson, J.D., N.H. Hopkins, J.W. Roberts, J.A. Steitz, and A.M. Weiner. 1987. Molecular biology of the gene. 4th ed. Benjamin/Cummings. Menlo Park, California. 1163 pp.

Watson, J.D., J. Tooze, and D.T. Kurtz. 1983. Recombinant DNA, A short course. Scientific American. Freeman. New York. 260 pp.

Wendel, J.F. and C.R. Parks. 1982. Genetic control of isozyme variation in *Camellia japonica* L. J. Hered. 73:197-204.

White, R. and J-M. Lalouel. 1988a. Sets of linked genetic markers for human chromosomes. Annu. Rev. Genet. 22:259-279.

White, R. and J-M. Lalouel. 1988b. Chromosome mapping with DNA markers. Sci. Amer. 258(2):40-48.

White, R. W., M. Leppert, D.T. Bishop, D. Barker, J. Berkowitz, C. Brown, P. Callahan, T. Holm, and L. Jerominski. 1985. Construction of linkage maps with DNA markers for human chromosomes. Nature 313:101-105.

Wolfe, S. L. 1983. Introduction to Cell Biology. Wadsworth, Inc. Elmont, Calif. 408 pp.

Wu, L., A.H. Harivandi, J.A. Harding, and W.B. Davis. 1984. Identification of Kentucky bluegrass cultivars with esterase and phosphoglucomutase isoenzyme markers. Crop Science 24:763-768.

Yoshida, T. 1985. Inheritance of susceptibility to *Citrus* Tristeza virus in trifoliate orange [*Poncirus trifoliata* Raf.]. (in Japanese). Bull. Fruit Tree Res. Stn. B 12:17-25.

Zamir, D. and S.D. Tanksley. 1988. Tomato genome is comprised largely of fast-evolving, low copy-number sequences. Mol. Gen. Genet. 213:254-261.

Zamir, D., S.D. Tanksley, and R.A. Jones. 1982. Haploid selection for low temperature tolerance of tomato pollen. Genetics 101:129-137.

BIOGRAPHIC SKETCH

Pan-chi Liou was born in Taiwan, Republic of China, on March 10, 1950. He attended the Agronomy Department in the National Taiwan University (Taipei, Taiwan) beginning in September, 1970; changed to the Horticultural Department in the same university in 1972; and received his Bachelor of Agricultural Science degree in June, 1974.

After two years of military service, he enrolled again in the Horticultural Department at the National Taiwan University in September, 1976, where he studied plant nutrition. He received his Master of Agricultural Science degree in June, 1978.

In 1979, he was employed as a Research Asistant in the Taiwan Agricultural Research Institute, Taichung, Taiwan, where he was mainly working for a *Citrus* budwood program and *Citrus* extension for the island. In 1982, sponsored by his government, he visited the U.S. for nine months to learn the *Citrus* budwood programs and *Citrus* nursery systems in California, Texas, and Florida.

In January, 1987, sponsored again by his government, he enrolled in the Fruit Crops Department at the University of Florida for his Ph.D. program, working in the area of *Citrus* genetics and plant molecular biology. Following completion of doctoral studies, the author will return to his country to work in the area of *Citrus* cultivar improvement.

The author and his wife, Li-hwa, are parents of a 10-year-old son, Jim, and a 4-year-old daughter, Ann.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gloria A. Moore

Gloria A. Moore, Chairman
Associate Professor of Horticultural Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Fred G. Gmitter

Fred G. Gmitter, Cochairman
Assistant Professor of Horticultural Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Christine D. Chase

Christine D. Chase
Assistant Professor of Horticultural Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jude W. Grosser

Jude W. Grosser
Associate Professor of Horticultural Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Paul Lyrene

Paul M. Lyrene
Professor of Horticultural Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

K. S. Derrick

Ken S. Derrick
Professor of Plant Pathology

This dissertation was submitted to the Graduate Faculty of the College of Agriculture and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1990

Jack L. Fry

Dean, College of Agriculture

Dean, Graduate School